

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:28:00 ; Search time 44 Seconds
(without alignments)
457,120 Million cell updates/sec

Title: US-09-705-911-24
Perfect score: 1401

Sequence: 1 MRAVFSTIIIPFNAAVTIQ.....ALTRLRKAVKSKERSAP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

- Database : Published Applications_AA.*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1 | 179 | 12.8 | 278 | US-09-767-041-36 | Sequence 36, Appl |
| 2 | 178 | 12.7 | 332 | US-09-767-041-22 | Sequence 22, Appl |
| 3 | 172 | 12.3 | 358 | US-09-815-242-5714 | Sequence 5714, Ap |
| 4 | 172 | 12.3 | 573 | US-09-815-242-12474 | Sequence 12474, A |
| 5 | 165.5 | 11.8 | 150 | US-09-924-358-29 | Sequence 29, Appl |
| 6 | 165 | 11.8 | 251 | US-09-738-626-4202 | Sequence 4202, Ap |
| 7 | 163 | 11.6 | 332 | US-09-767-041-21 | Sequence 21, Appl |
| 8 | 162.5 | 11.6 | 274 | US-09-738-626-3905 | Sequence 3905, Ap |
| 9 | 159.5 | 11.4 | 120 | US-09-767-041-51 | Sequence 51, Appl |
| 10 | 158 | 11.3 | 348 | US-10-007-267-11 | Sequence 11, Appl |
| 11 | 158 | 11.3 | 348 | US-10-007-267-3 | Sequence 3, Appl |
| 12 | 157.5 | 11.2 | 706 | US-09-815-242-4950 | Sequence 4950, Ap |
| 13 | 157.5 | 11.2 | 715 | US-09-815-242-4950 | Sequence 4950, Ap |
| 14 | 156.5 | 11.2 | 270 | US-09-815-242-39 | Sequence 39, Appl |
| 15 | 155.5 | 11.1 | 120 | US-09-767-041-52 | Sequence 52, Appl |
| 16 | 153 | 10.9 | 322 | US-09-767-041-34 | Sequence 34, Appl |
| 17 | 152 | 10.8 | 322 | US-09-767-041-35 | Sequence 35, Appl |
| 18 | 150 | 10.7 | 187 | US-09-973-457-4 | Sequence 4, Appl |
| 19 | 150 | 10.7 | 187 | US-09-815-028-7 | Sequence 7, Appl |

| | | | | | |
|----|-------|------|-----|--------------------|--------------------|
| 20 | 150 | 10.7 | 187 | US-10-074-527-4 | Sequence 4, Appl |
| 21 | 149.5 | 10.7 | 972 | US-09-879-959-10 | Sequence 10, Appl |
| 22 | 146.5 | 10.5 | 301 | US-09-816-028A-27 | Sequence 27, Appl |
| 23 | 145.5 | 10.4 | 345 | US-09-738-626-3894 | Sequence 3894, Ap |
| 24 | 144.5 | 10.3 | 281 | US-09-765-272-196 | Sequence 196, Appl |
| 25 | 144.5 | 10.3 | 297 | US-09-816-028A-31 | Sequence 31, Appl |
| 26 | 141 | 10.1 | 303 | US-09-816-028A-29 | Sequence 29, Appl |
| 27 | 134.5 | 9.6 | 674 | US-09-765-272-200 | Sequence 200, Appl |
| 28 | 132 | 9.4 | 243 | US-09-767-041-47 | Sequence 47, Appl |
| 29 | 131 | 9.4 | 210 | US-09-738-626-4733 | Sequence 4733, Ap |
| 30 | 129 | 9.2 | 389 | US-09-816-028A-34 | Sequence 34, Appl |
| 31 | 128 | 9.1 | 337 | US-10-007-267-12 | Sequence 12, Appl |
| 32 | 128 | 9.1 | 337 | US-10-007-267-5 | Sequence 5, Appl |
| 33 | 123.5 | 8.8 | 418 | US-09-816-028A-33 | Sequence 33, Appl |
| 34 | 108.5 | 7.7 | 270 | US-09-765-272-198 | Sequence 198, Appl |
| 35 | 108.5 | 7.7 | 303 | US-09-765-272-202 | Sequence 202, Appl |
| 36 | 107.5 | 7.7 | 272 | US-09-934-899-12 | Sequence 12, Appl |
| 37 | 107.5 | 7.7 | 272 | US-09-934-899-12 | Sequence 12, Appl |
| 38 | 105 | 7.5 | 313 | US-09-934-899-12 | Sequence 12, Appl |
| 39 | 102 | 7.3 | 440 | US-09-902-525-46 | Sequence 1, Appl |
| 40 | 100 | 7.1 | 317 | US-09-934-899-18 | Sequence 18, Appl |
| 41 | 100 | 7.1 | 317 | US-09-934-899-18 | Sequence 18, Appl |
| 42 | 96.5 | 6.9 | 395 | US-10-042-523-4 | Sequence 38, Appl |
| 43 | 94.5 | 6.7 | 398 | US-09-741-669-430 | Sequence 430, Appl |
| 44 | 92.5 | 6.6 | 256 | US-09-925-301-883 | Sequence 883, Appl |
| 45 | 92.5 | 6.6 | 269 | US-09-767-041-41 | Sequence 41, Appl |

ALIGNMENTS

RESULT 1
US-09-767-041-36
Sequence 36, Application US/09767041
Patent No. US20020055168A1
GENERAL INFORMATION:
APPLICANT: Smith, Hilda
TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
FILE REFERENCE: 2183-426
CURRENT APPLICATION NUMBER: US/09/767, 041
CURRENT FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: PCF/NL99/00460
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: EP8202465.5
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: EP98202467.1
PRIOR FILING DATE: 1998-07-22
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.0
SEQ ID NO 36
LENGTH: 278
TYPE: PRT
ORGANISM: Streptococcus suis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: CP51K
US-09-767-041-36
Query Match 12.8%; Score 179; DB 10; Length 278;
Best local Similarity 23.6%; Pred. No. 3.6e-11;
Matches 61; Conservative 47; Mismatches 93; Indels 58; Gaps 10;
DB 2 TAVVFSIIIPFNAAVTIQACLSIVGQYREVEVLVDGSDTRTIDIANSPFELGSR 61
DB 3 TISKISIIIVINVEKYLKSIDISIVNQYKHEILLVNGSTDNSEICLAAYAKK-DSR 61
DB 62 LVVHSGDDPVPYAMNNGVAVGCEWFLFGADDTL-----EPTIAQVAAF 109
DB 62 IIRFKKENGSLSPKRNIGISRAKGDYLAFTDSDFHSEFIQRLHEALIERENMLVNAVGY 121
DB 110 LGDHAASHLYGVGVMSKSRNAGPFDLRLLFETNLC-----HOSI-----FYRR 156
DB 122 DRVDASGHFLTAFLPIIN-----QAVLSGRVCKKLEADGHRVAVACNKLYKK 170

OY 157 ELFDGIGYNYLR---RWADMDENIRC---FSNPALITRYMDVVISYNDMTGFSMRG 210
 Db 171 ELFE-----DEFREKGIHEDEYFYRLYLEKVAIVECLYYVDRENSITSSM--- 222
 OY 211 TDKEF-----RKRLPMY 222
 Db 223 TDHREHCLLEFONERMDRY 241

RESULT 2

US-09-767-041-22
 ; Sequence 22, Application US/09767041
 ; Patent No. US2002005168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Hilda
 ; TITLE OF INVENTION: STREPTOCOCCUS SUTS VACCINES AND DIAGNOSTIC TESTS
 ; FILE REFERENCE: 2183-4726
 ; CURRENT APPLICATION NUMBER: US/09/767,041
 ; CURRENT FILING DATE: 2001-01-22
 ; PRIOR APPLICATION NUMBER: PCT/NL99/00460
 ; PRIOR FILING DATE: 1999-07-19
 ; PRIOR APPLICATION NUMBER: EP98202465.5
 ; PRIOR FILING DATE: 1998-07-22
 ; PRIOR APPLICATION NUMBER: EP98202467.1
 ; PRIOR FILING DATE: 1998-07-22
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: Patent version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 332
 ; TYPE: PRN
 ; ORGANISM: Streptococcus suis
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: CPS2K
 US-09-767-041-22

Query Match 12.7%, Score 178; DB 10; Length 332;
 Best Local Similarity 21.9%, Pred. No. 5.9e-11;
 Matches 68; Conservative 53; Mismatches 100; Indels 90; Gaps 12;

OY 7 STIITFNAAVTLQACLSIVGQTYREVYVAVDGSITRTIDIANSPRELGSRLVMS 66
 Db 5 STIIVYVEQYLSKISINVTQYKHEILLVNGSTDNSEICLAYAKK-DSRIRYRK 63
 OY 67 GPDGPPYAMNNGVAVGAVFLGADDTLY-----EPTTLAQAFAFGDHA 114
 Db 64 KENGSLSPARNYGISRAKGDYLAFLDSDFIHSFIOQLHEALIERENALVAVAGYDRVDA 123
 OY 115 ASHLVYGVVMSSTKSRHAGPPDLDLLEFETNLC-----HOSI-----FYRRELFDG 161
 Db 124 SGHFLTAEPFLPTN-----OAVLSGRNVCKKLEADGHRFVAVANNKLYKKELFD- 171
 OY 162 IGVNLYAR---RVADMDENIRC---FSNPALITRYMDVVISYNDMTGFSMRGTDKEF 215
 Db 172 ---FREKGIHEDEYFYRLYLEKVAIVECLYYVDRENSITSSM---TDHRE 223
 OY 216 ---RKRLPMY-----FWVGMETCRRLM---AF 227
 Db 224 HCLLEFONERMDRYESRGDKELLLECYNSFLAFVFLGKYNHWSKOOKKLOTLEIRIY 283
 OY 238 LKDKENRRIAL 248
 Db 284 KOLKONKRIAL 294

RESULT 3
 US-09-815-242-5714
 ; Sequence 5714, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlssen, Karl L.

APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5714
 ; LENGTH: 358
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5714

Query Match 12.3%, Score 172; DB 10; Length 358;
 Best Local Similarity 33.1%, Pred. No. 2.9e-10;
 Matches 51; Conservative 25; Mismatches 66; Indels 12; Gaps 5;

OY 6 FSIIITFNAAVTLQACLSIVGQTY--REVEVAVDGSITRTIDIANSPRELGS---S 60
 Db 1 FSVITVYVSEKITYTELNSLANAKQDFPKTEFEVYVVDGSDTQIQIVEYRKMLNKYS 60
 OY 61 RLTVHSGPDGPPYAMNNGVAVGAVFLGADDTLYEPTTLAQAFAFGDHAASHLY 120
 Db 61 QLEITNSG---GPKCPRVVAVLKAQAEVEFLVDSDDYI-NKETLKDAAFIDEHRSDVLL- 115
 OY 121 GDVVMSTKSRHAGPPDLDLLEFETNLC-HOSITY 154
 Db 116 --TKMGVNGRGVPOSMFKETADEVTLLNSRITY 147

RESULT 4
 US-09-815-242-12474
 ; Sequence 12474, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlssen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12474
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12474

```

```

Query Match 12.3%; Score 172; DB 10; Length 573;
Best Local Similarity 33.1%; Pred. No. 5.5e-10;
Matches 51; Conservative 25; Mismatches 66; Indels 12; Gaps 5;

```

```

QY 6 FSIITPTNAATLQACISIVGQY--REVEVVLVDGSGTPTDIDIANSPRELG---S 60
  ||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 4 FSVIPTVSEKYTEILNLSLAKQDPKTEFEVYVDDSTQITQIVERYKMLKYS 63
QY 61 RLVHSGPDDGPDYDAMNRCGVATGEMVFLGADDTLYEPTTLAQVAFLGDHAAHLVY 120
  : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 64 QLETNSG---GPKPRNVALKQAGEFVLFVDSDDYI--NKEILKDAALFIDEHSDVLL- 118
QY 121 GCVVARSSTRHAGPDDRLFLFTNLCHQSIFY 154
DB 119 --IKKGVNGRVQPMFKETAPEVTLNLSRIY 150

```

```

RESULT 5
US-09-924-358-29
; Sequence 29, Application US/09924358
; Patent No. US20020107376A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Macbeth, Kyle
; TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
; TITLE OF INVENTION: 58764,
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE FAMILY MEMBERS AND USES THEREFOR
; FILE REFERENCE: 38155-20034.00
; CURRENT APPLICATION NUMBER: US/09/924,358
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/229,300
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid
US-09-924-358-29

```

```

Query Match 11.8%; Score 165.5; DB 10; Length 150;
Best Local Similarity 35.0%; Pred. No. 4.3e-10;
Matches 55; Conservative 26; Mismatches 63; Indels 13; Gaps 8;
QY 7 SIIITPTNAATLQACISIVGQY--TYREVEVVLVDGSGTPTDIDIANSPRELGLVYH 65
  ||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 SIIITPTNKKYLECLESILNMQTYENELIVVDGSGTGEVLEIEYAKPDIR-VIR 59
QY 66 SPPDDGPDYDAMNRCGVATGEMVFLGADDTLYEPTTLAQVAFLGDHAAHLVYGY 123
DB 60 LEENIGLAAARNAAGLKHATGVDYIAFLDADDEV--PDLKELLELEKNGAD-ITYGRV 116
QY 124 VNRSTRHAGPDDRLFLFTNLCHQSIFYRRELE 160

```

```

DB 117 I-NENKGRINGKLRL--LVF---LIGSNALYRREAL 147

```

```

RESULT 6
US-09-738-626-4202
; Sequence 4202, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SETKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4202
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4202

```

```

Query Match 11.8%; Score 165; DB 9; Length 251;
Best Local Similarity 27.4%; Pred. No. 1e-09;
Matches 76; Conservative 43; Mismatches 100; Indels 58; Gaps 15;
QY 1 MTAAPV---FSIITPTNAATLQACISIVGQYREVEVVLVDGSGTPTDIDIANSPRE 57
  ||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MTSIORISITVPCANDAVLREGCLSSIVAQITQPEVIVVDGSGTSDNSVEVAN----R 56
QY 58 LGSRLVYHSGPDDGPDYDAMNRCGVATGEMVFLGADDTLYEPTTLAQVAFLGDHAAH 117
  : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 57 MGAR-VVHE--PLOGITWASAGYNSANGDLIVRDA--DCVYIPDHLQVNAIWN---RTE 110
QY 118 LVYGDVVMRSTRHAGPDDRLFLFE--TNLCHQSIFYRRELEFGIGPPYNLRYVMADWD 176
DB 111 ETEGRTVALTGT---GSFPIGRMGWALCTYIGA--YRNSTKALHPTI----- 157
QY 177 ENIRCESNPALITRYM-----DVVISE---YNDM-TGFSNRQGTDEFRRLPMYF--- 223
DB 158 -----FGSNSVIRQWMDVYKQDITLSEFVHEMDYSFVRPHETVWEKNTLIMPR 212
QY 224 -----WAGWETCRMAFLKDKENRLALR 249
DB 213 ALMGVRSFIRLVGFYTIK--IWKREPVHQRLESR 247

```



```

1 SOFTWARE: PatentIn Release #1.0, Version #1.25
2
3 CURRENT APPLICATION DATA:
4   APPLICATION NUMBER: US/10/007,267
5   FILING DATE: 03-Dec-2001
6   CLASSIFICATION: <Unknown>
7
8 PRIOR APPLICATION DATA:
9   APPLICATION NUMBER: US/09/333,412
10  FILING DATE: 15-Jun-1999
11  APPLICATION NUMBER: 08/312,387
12  FILING DATE: July 7, 1994
13  ATTORNEY/AGENT INFORMATION:
14    NAME: Jackson Esq., David A.
15    REGISTRATION NUMBER: 26,742
16    REFERENCE/DOCKET NUMBER: 600-1-095
17  TELECOMMUNICATION INFORMATION:
18    TELEPHONE: 201 487-5800
19    TELEFAX: 201 343-1684
20    TELEX: 133521
21
22 INFORMATION FOR SEQ ID NO: 11:
23   SEQUENCE CHARACTERISTICS:
24     LENGTH: 348 amino acids
25     TYPE: amino acid
26     TOPOLOGY: 1linear
27   MOLECULE TYPE: protein
28   SEQUENCE DESCRIPTION: SEQ ID NO: 11:
29
30 US-10-007-267-11
31
32 Query Match      11.3%; Score 158; DB 12; Length 348;
33 Best Local Similarity 23.7%; Pred. No. 8.9e-09;
34 Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
95
```

```

? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 4950
? LENGTH: 706
? TYPE: PRT
? ORGANISM: Enterococcus faecalis
? DS-09-815-242-4950

```

| | | | | |
|--------------------------|-------|-------------------|-----------|------------|
| Query Match | 11.2% | Score 157.5 | DB 10 | length 706 |
| Best Local Similarity | 26.1% | Pred. No. 2.7e-08 | | |
| Matches 61, Conservative | 30 | Mismatches 98 | Indels 45 | Gaps 8 |

```

QY 4 PVESIIPTFENNAAT--LQACIGSVGTQVEVEVVLVDGSGSTRPTDIANSFPELGSRL 62
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  180 PKISIAIPVYVVEEKWLRICIDSLNGLVYINWELCMADDASTDPNWKILTEQGLDERI 239
-1 63 VYHSGPDDGPY-DAMNRGVGVAITEWVLTGADDTL-----YEPTTLAOVAAFLGDHAAS 116
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 240 RIVERQNGIHISEATNSALMATGEFALLDNDELAINAFYE-----VYKVLNENPEL 293
QY 117 HLTVGD---VYMRSTKSRHAGPEDLD-RLLEFNLCHQSIFYRRELFDGIPYNLARYRW 172
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 DLIVSDEDKIDMDGNRSRDPAFKPMWSDDLITGLTYISHLGVYRSRIIEEIGCGRRGEGS 353
QY 173 ADMQFNIRCSNPALITRRMDVYISEVNDMTGFSMRGCTGKERKRLRP--MYRW 224
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 354 QDYDLVLR-----ETKTKTKERTITHPKVLYYW 381

```

RESULT 13
US-09-815-242-10511
Sequence 10511, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10511
LENGTH: 715
TYPE: PRT

ORGANISM: *Enterococcus faecalis*
US-09-815-242-10511

| | | | | |
|-----------------------|--------|--------------------|----------------|-------------------|
| Query Match | 11.2%; | Score 157.5; | DB 10; | Length 715; |
| Best Local Similarity | 26.1%; | Pred. No. 2.7e-08; | | |
| Matches | 61; | Conservative 30; | Mismatches 98; | Indels 45; Gaps 8 |

```
QY      4 PVFSLIPTFNAAAT--LOACIGSVIVGTQREVEVALVDGSTDFTLLANSFRELCSRL 62
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     182 PKSIAMVYVNEEKWLRCTCIDSTLNLQYTYMWECLMDADSTDPNVKKILTEYQQDDBRI 244
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      63 VVHSGPDDGPY-DAMNRGVGVATGEWWLFGLGADDTL-----YEFTTLQVAAPFLGDHAAS 116
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     242 RVEREONGHSISEATNSALAIATAGEFALLDNDEDLAINAFYE-----YVKVLNENDEL 295
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      117 HLAVGD---VMRSTKSHAGPFOLD-RLLEFWMLCHQSIFRYRELFDGICPNLRIRWV 172
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     296 DLITSDECKIDMDNRSRDPARPKPMSPDLLLTGYISHLGAYRRSILEETGGFKEGGS 355
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      173 ADMFENCISFPNALLITRYMDVIVSEYNDMTGFSMRQGTOKFEFRKLRP--MYEM 224
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     356 QDIYDLVLR-----FTEKTTRKERITTHLPKULYYW 383
```

```

1 RESULT 14
2 US-09-816-028A-39
3 Sequence 39, Application US/09816028A
4 Patent No. US20020042369A1
5 GENERAL INFORMATION:
6 APPLICANT: Gilbert, Michel
7 APPLICANT: Makarchuk, Warren W.
8 APPLICANT: National Research Council of Canada
9 TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
10 TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
11 FILE REFERENCE: 019633-000111US
12 CURRENT APPLICATION NUMBER: US/09/816,028A
13 CURRENT FILING DATE: 2001-03-21
14 PRIOR APPLICATION NUMBER: US 60/118,213
15 PRIOR FILING DATE: 1999-02-01
16 PRIOR APPLICATION NUMBER: US 09/495,406
17 PRIOR FILING DATE: 2000-01-31
18 NUMBER OF SEQ ID NOS: 49
19 SOFTWARE: PatentIn Ver. 2.1
20 SEQ ID NO 39
21 LENGTH: 270
22 TYPE: PR1
23 ORGANISM: Campylobacter jejuni
24 FEATURE:
25 OTHER INFORMATION: glycosyltransferase from C. jejuni OH4384 (ORF 12a
26 US-09-816-028A-39
27 OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)

```

[illegible]

RESULT 15

US-09-767-041-52

; Sequence 52, Application US/09767041
; Patent No. US20020055168A1

GENERAL INFORMATION:

; APPLICANT: Smith, Hilda

; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS

; FILE REFERENCE: 2183-4726

; CURRENT APPLICATION NUMBER: US/09/767,041

; PCT/US99/00460

; PRIOR FILING DATE: 1999-07-19

; PRIOR APPLICATION NUMBER: EP98202465.5

; PRIOR FILING DATE: 1998-07-22

; PRIOR APPLICATION NUMBER: EP98202467.1

; NUMBER OF SEQ ID NOS: 53

; SOFTWARE: PatentIn version 3.0

SEQ ID NO 52

LENGTH: 120

; TYPE: PRT

; ORGANISM: Streptococcus suis

FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: N-terminal part of CPS2K

; NAME/KEY: misc_feature

; LOCATION: (1)..(120)

; OTHER INFORMATION: Xaa may be any amino acid

US-09-767-041-52

Query Match 11.1%; Score 155.5; DB 10; Length 120;

Best Local Similarity 34.8%; Pred. No. 3,8e-09;

Matches 32; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

OY 7 SIIPFNAATVLTQACISVGTREVEVLVWGSGTDTLIDIANSPFELGSRPLVHS 66

DB 5 SIIVPIYNEOYLKSCINISIVNQYKHIELVNDGSSITDENSEICLAIYAKK-DSRIIRYFK 63

OY 67 GPDDGPFYDAMNRGVATGEWVFLGADDTLY 98

DB 64 KENGGLSDARNYGISRAKGDYLAFLIDSDDFIH 95

Search completed: April 17, 2003, 17:36:37
Job time : 46 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:25:41 ; Search time 29 Seconds
(Without alignments) 269.879 Million cell updates/sec

Title: US-09-705-911-24

Sequence: 1 MTAPVFSIIIPFNNAVTLQ.....ALTRILIRKAVKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Marched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/pdata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/pdata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/pdata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/pdata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 161 | 11.5 | 324 | 1 | US-08-597-236-10 |
| 2 | 161 | 11.5 | 324 | 1 | US-08-746-682A-10 |
| 3 | 158 | 11.3 | 348 | 1 | US-08-312-387B-3 |
| 4 | 158 | 11.3 | 348 | 1 | US-08-312-387B-11 |
| 5 | 158 | 11.3 | 348 | 1 | US-08-683-426-3 |
| 6 | 158 | 11.3 | 348 | 1 | US-08-683-426-11 |
| 7 | 158 | 11.3 | 348 | 1 | US-08-683-426-3 |
| 8 | 158 | 11.3 | 348 | 1 | US-08-683-426-11 |
| 9 | 158 | 11.3 | 348 | 1 | US-08-683-426-3 |
| 10 | 158 | 11.3 | 348 | 1 | US-08-683-426-11 |
| 11 | 158 | 11.3 | 348 | 2 | US-08-878-360-3 |
| 12 | 158 | 11.3 | 348 | 2 | US-08-878-360-11 |
| 13 | 158 | 11.3 | 348 | 3 | US-08-478-140B-8 |
| 14 | 158 | 11.3 | 348 | 4 | US-08-478-140B-8 |
| 15 | 158 | 11.3 | 348 | 4 | US-09-333-412-3 |
| 16 | 158 | 11.3 | 348 | 4 | US-09-333-412-11 |
| 17 | 147.5 | 10.5 | 281 | 4 | US-09-338-943-3 |
| 18 | 147.5 | 10.5 | 281 | 4 | US-09-338-943-8 |
| 19 | 144.5 | 10.3 | 281 | 4 | US-09-437-277-3 |
| 20 | 141 | 10.1 | 702 | 4 | US-08-961-083-196 |
| 21 | 135.5 | 9.7 | 93 | 4 | US-08-961-083-196 |
| 22 | 128 | 9.1 | 674 | 4 | US-08-858-207A-521 |
| 23 | 128 | 9.1 | 337 | 1 | US-08-312-387B-5 |
| 24 | 128 | 9.1 | 337 | 1 | US-08-312-387B-12 |
| 25 | 128 | 9.1 | 337 | 1 | US-08-683-426-5 |
| 26 | 128 | 9.1 | 337 | 1 | US-08-683-426-12 |
| 27 | 128 | 9.1 | 337 | 1 | US-08-683-426-5 |

| | | | | | | |
|----|-------|-----|------|---|---------------------|--------------------|
| 28 | 128 | 9.1 | 337 | 2 | US-08-878-360-5 | Sequence 5, Appl1 |
| 29 | 128 | 9.1 | 337 | 2 | US-08-878-360-12 | Sequence 12, Appl1 |
| 30 | 128 | 9.1 | 337 | 3 | US-08-478-140B-5 | Sequence 5, Appl1 |
| 31 | 128 | 9.1 | 337 | 4 | US-09-333-412-5 | Sequence 5, Appl1 |
| 32 | 128 | 9.1 | 337 | 4 | US-09-333-412-12 | Sequence 12, Appl1 |
| 33 | 128 | 9.1 | 337 | 4 | US-09-338-943-5 | Sequence 5, Appl1 |
| 34 | 115.5 | 8.2 | 418 | 4 | US-09-134-001C-4051 | Sequence 4051, Ap |
| 35 | 108.5 | 7.7 | 270 | 4 | US-08-961-083-198 | Sequence 198, Ap |
| 36 | 108.5 | 7.7 | 303 | 4 | US-08-961-083-202 | Sequence 202, Ap |
| 37 | 99 | 7.1 | 358 | 4 | US-09-134-001C-5633 | Sequence 5633, Ap |
| 38 | 96.5 | 6.9 | 395 | 4 | US-08-635-552A-4 | Sequence 4, Appl1 |
| 39 | 96.5 | 6.9 | 419 | 2 | US-08-270-581-2 | Sequence 2, Appl1 |
| 40 | 96.5 | 6.9 | 419 | 4 | US-09-146-893-2 | Sequence 2, Appl1 |
| 41 | 92 | 6.6 | 727 | 4 | US-09-134-001C-4067 | Sequence 4067, Ap |
| 42 | 84.5 | 6.0 | 79 | 4 | US-08-961-083-168 | Sequence 168, Ap |
| 43 | 76 | 5.4 | 723 | 4 | US-08-960-048-9 | Sequence 9, Appl1 |
| 44 | 76 | 5.4 | 723 | 5 | PCT-US91-01726-4 | Sequence 4, Appl1 |
| 45 | 76 | 5.4 | 3724 | 2 | US-08-804-227C-10 | Sequence 10, Appl1 |

ALIGNMENTS

RESULT 1
US-08-597-236-10
Sequence 10, Application US/08597236
Patent No. 5733765
GENERAL INFORMATION:
APPLICANT: STINGELE, Francesca
APPLICANT: MOLETT, Beat
TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
TITLE OF INVENTION: EXOPOLYSACCHARIDES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edwards
STREET: 1155 Avenue of the Americans
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/597,236
FILING DATE:
CLASSIFICATION: 426
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-597-236-10

Query Match 11.5%; Score 161; DB 1; Length 324;
Best Local Similarity 35.9%; Pred. No. 2e-10;

Matches 42; Conservative 17; Mismatches 44; Indels 14; Gaps 3;

QY 5 VFIIIPFNNAVTLQACISIVGQTYREVEVAVDGSRTDRIANSPFELGSRVYV 64
DB 6 LIIIVPVVVEVLEICLOSVOQNTNNREVIIVNDGSDSSLSICEKFNQ-DKRSV 64
QY 65 HSPDDGPPIDAMNRGVAVNGEWTFLGADDTLEPTTLQVAAFLGDSHAASHLYVG 121

DB 65 FSKENGSSARNFGIKAKGSFTTFVDSDDYI-----VKDYL-----SHLVAG 108

RESULT 2

US-08-746-682A-10
Sequence 10, Application US/08746682A

Patent No. 5786184

GENERAL INFORMATION:

APPLICANT: STINGELE, Francesca

APPLICANT: MOLETT, Beat

TITLE OF INVENTION: LACTIC BACTERIA PRODUCING

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americans

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/746,682A

FILING DATE: 14-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/597,236

FILING DATE: 20-JUN-1995

APPLICATION NUMBER: EP 95201669.9

FILING DATE: 20-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fanucci A., Allan

REGISTRATION NUMBER: 30256

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 324 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-746-682A-10

Query Match 11.5%; Score 161; DB 1; Length 324;

Best Local Similarity 35.9%; Pred. No. 2e-10; Matches 42; Conservative 17; Mismatches 44; Indels 14; Gaps 3;

5 VESIIIPFENAAVTTLQACLSIVGQTYREVEVVLVDGSGTDRITLDIANSFPELGSRLV 64

6 LLSIYIPVNVKYLEKCSQVQNGYNNFEVLVDGSDTSLSCFKNQ-DKRFV 64

QY 65 HSGPDDGPDAMNRGVAVTGEMVFLGADDTLYEPTTLAQVAALGDHAASHLVY 121

DB 65 FSKENGSSARNFGIKAKGSFTTFVDSDDYI-----VKDYL-----SHLVAG 108

RESULT 3

US-08-312-387B-3

Sequence 3, Application US/08312387B

Patent No. 5545553

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/312,387B

FILING DATE: July 7, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 348 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-312-387B-3

Query Match 11.3%; Score 158; DB 1; Length 348;

Best Local Similarity 23.7%; Pred. No. 4.9e-10; Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY 4 PVSIIIPFENAAVTTLQACLSIVGQTYREVEVVLVDGSGTDRITLDIANSFPELGSRLV 63

DB 3 PLSVILICAVNVEKYFAQSLAAVNGTWRNLDILYVDGSDTGLTALAKDFOKDSRIKI 62

QY 64 VHSQPDGPDAMNRGVAVTGEMVFLGADDTLYEPTTLAQVAALGDHAASHLVY 105

DB 63 LQAQNSGLIPSLINIGDELAKSGGGGEXIARTDADDIASPCWIEKIVGEMEDRSIIA 122

QY 106 VAAFL-----GDHAASHLVYGDVYMRSTKSRHAGFPDLRLFPETNLCHOSIFRRE 157

DB 123 MGAWEVLSERKQGNRLARHKHKIKKPTRHEDIAF-----PFGNPHTNNTIMRRS 178

QY 158 LFDGIPYNLRYRWADW 175

DB 179 VIDG-----GLRYPDERDW 192

RESULT 4

US-08-312-387B-11

Sequence 11, Application US/08312387B

Patent No. 5545553

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/312,387B

FILING DATE: July 7, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

[illegible]

DB 123 MGAWLEVLSEERKGNRLAHKHKIKMKPTRHEDIAAF-----PFGNPIHNNMTIMRRS 178
QY 158 LFDGIGPYNLRYRWADW 175
DB 179 VIDG-----GLRYDTERDW 192

RESULT 9
US-08-878-360-3
; Sequence 3, Application US/08878360
; Patent No. 5945322

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-3

Query Match 11.3%; Score 158; DB 2; Length 348;
Best Local Similarity 23.7%; Pred. No. 4.9e-10;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY 4 PVFSIIIFPENAAYTLQACISIVGQTYREVEVVLVDGSGTDRDLDIANSFRPELGSRV 63
DB 3 PLVSVLICAVNVEKYFAQSLAAVNOGTWRNLDILIVDGSSTGTALAKDFQRDSRIKI 62
QY 64 VHSPPDDGPDYDAMRGV-----GVATGEWVLFAGD-----TLYEPTTLAQ 105
DB 63 LAQAQNSGLIPSLINIGDLBLAKSGGGGGGYIARTDADDIASPCWIKIYVEMKDSITA 122
QY 106 VAAFL-----GDHAAASHLYVGDVYMRSTKSRHAGFPDLRLLEFNTLCHOSIFYRRE 157
DB 123 MGAWLEVLSEERKGNRLAHKHKIKMKPTRHEDIAAF-----PFGNPIHNNMTIMRRS 178
QY 158 LFDGIGPYNLRYRWADW 175
DB 179 VIDG-----GLRYDTERDW 192

DB 179 VIDG-----GLRYDTERDW 192

RESULT 10
US-08-878-360-11
; Sequence 11, Application US/08878360
; Patent No. 5945322

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-11

Query Match 11.3%; Score 158; DB 2; Length 348;
Best Local Similarity 23.7%; Pred. No. 4.9e-10;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY 4 PVFSIIIFPENAAYTLQACISIVGQTYREVEVVLVDGSGTDRDLDIANSFRPELGSRV 63
DB 3 PLVSVLICAVNVEKYFAQSLAAVNOGTWRNLDILIVDGSSTGTALAKDFQRDSRIKI 62
QY 64 VHSPPDDGPDYDAMRGV-----GVATGEWVLFAGD-----TLYEPTTLAQ 105
DB 63 LAQAQNSGLIPSLINIGDLBLAKSGGGGGGYIARTDADDIASPCWIKIYVEMKDSITA 122
QY 106 VAAFL-----GDHAAASHLYVGDVYMRSTKSRHAGFPDLRLLEFNTLCHOSIFYRRE 157
DB 123 MGAWLEVLSEERKGNRLAHKHKIKMKPTRHEDIAAF-----PFGNPIHNNMTIMRRS 178
QY 158 LFDGIGPYNLRYRWADW 175
DB 179 VIDG-----GLRYDTERDW 192

RESULT 11
US-08-478-140B-3


```

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/478,140
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: 1linear
MOLECULE TYPE: protein
US-09-338-943-3

```

```

Query Match      11.3%; Score 158; DB 4; Length 348;
Best Local Similarity 23.7%; Pred. No. 4,9e-10;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

```

```

QY 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVVLVDGSGSTDRTIDIANSPREPGLSRLV 63
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3 PLVSVLICAVNYEKYFAQSLAIVNQTWRNLDILIVDGSSTGTLAIANDFOKDRSRIKI 62
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 64 VHSQDDGPTDAMNGV-----GVATGEWVLEFLGAD-----TLYEPTTLAQ 105
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 63 LQAQNSGLIPSLINIGLELAKSGGGGGEYIARTDADIASPGWIEKIVGEMEKDRSIIA 122
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 106 VAAFL-----GDAASHLVYGDVVMRSTKSRHAGPFIDLRLAFETNLCQHSIFYRRE 157
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 123 MGAWLEVLSEEDGNRIAHKHKIKWKPTRHEDIAAF---FPGNPITHNTMTMRRS 178
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 158 LFDGIGPYNLRRYVADW 175
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 179 VIDG---GLRYDTERDW 192
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Search completed: April 17, 2003, 17:29:27
Job time : 30 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:23:35 ; Search time 33 Seconds

(without alignments)
1660.865 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401

Sequence: 1 MFAPVFSIIIPFNAAVTLQ.....ALTRLIRKAVSKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Targeted: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organella:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 1401 | 100.0 | 266 | 2 | 088109 | 088109 mycobacteri |
| 2 | 298.5 | 21.3 | 297 | 2 | 093025 | 093025 salmonella |
| 3 | 298 | 21.3 | 297 | 2 | 093025 | 093025 salmonella |
| 4 | 276 | 19.7 | 265 | 2 | 08VWX0 | 08VWX0 leptospira |
| 5 | 275 | 19.6 | 265 | 2 | 09AEE4 | 09AEE4 leptospira |
| 6 | 275 | 19.6 | 265 | 2 | 08VWX7 | 08VWX7 leptospira |
| 7 | 275 | 19.6 | 265 | 2 | 08VWX5 | 08VWX5 leptospira |
| 8 | 274 | 19.6 | 265 | 2 | 08VWX9 | 08VWX9 leptospira |
| 9 | 271 | 19.3 | 265 | 2 | 08VWX4 | 08VWX4 leptospira |
| 10 | 267.5 | 19.1 | 248 | 2 | 08VWX1 | 08VWX1 leptospira |
| 11 | 267.5 | 19.1 | 248 | 2 | 08VWX8 | 08VWX8 leptospira |
| 12 | 266.5 | 19.0 | 248 | 2 | 09EXY4 | 09EXY4 leptospira |
| 13 | 255.5 | 18.2 | 251 | 2 | 09EXY3 | 09EXY3 leptospira |
| 14 | 255 | 18.2 | 254 | 2 | 09EXY2 | 09EXY2 leptospira |
| 15 | 254.5 | 18.2 | 263 | 2 | 09EXY1 | 09EXY1 leptospira |
| 16 | 252 | 18.0 | 254 | 2 | 09EXY9 | 09EXY9 salmonella |

| | | | | | | |
|----|-------|------|-----|----|--------|---------------------|
| 17 | 248 | 17.7 | 254 | 2 | 09XDJ3 | 09XDJ3 bacterioides |
| 18 | 247 | 17.6 | 276 | 2 | 09RQ94 | 09RQ94 rhizobium e |
| 19 | 238.5 | 17.0 | 247 | 2 | 09RCB0 | 09RCB0 yersinia ps |
| 20 | 238.5 | 17.0 | 247 | 16 | 09RCC4 | 09RCC4 yersinia pe |
| 21 | 232.5 | 16.6 | 260 | 16 | 09ZMX1 | 09ZMX1 helicobacte |
| 22 | 231.5 | 16.5 | 259 | 16 | 024928 | 024928 helicobacte |
| 23 | 220 | 15.7 | 324 | 16 | 08VSL1 | 08VSL1 anabaena sp |
| 24 | 207 | 14.8 | 321 | 16 | 08VSM2 | 08VSM2 anabaena sp |
| 25 | 207 | 14.8 | 368 | 16 | 098JH2 | 098JH2 rhizobium l |
| 26 | 206.5 | 14.7 | 316 | 16 | 08VSL6 | 08VSL6 anabaena sp |
| 27 | 205.5 | 14.7 | 298 | 17 | 09VZ16 | 09VZ16 pyrococcus |
| 28 | 204.5 | 14.6 | 330 | 16 | 08VSM1 | 08VSM1 anabaena sp |
| 29 | 201.5 | 14.4 | 321 | 2 | 09AFH3 | 09AFH3 streptococc |
| 30 | 200 | 14.3 | 314 | 2 | 09LAB8 | 09LAB8 aeromonas h |
| 31 | 196.5 | 14.0 | 337 | 16 | 08VXD6 | 08VXD6 anabaena sp |
| 32 | 196.5 | 14.0 | 367 | 16 | 092V61 | 092V61 rhizobium m |
| 33 | 192 | 13.7 | 336 | 16 | 097H39 | 097H39 clostridium |
| 34 | 190.5 | 13.6 | 323 | 16 | 08VSL7 | 08VSL7 anabaena sp |
| 35 | 190 | 13.6 | 248 | 16 | 08VSL7 | 08VSL7 anabaena sp |
| 36 | 189.5 | 13.5 | 333 | 16 | 08VW50 | 08VW50 escherichia |
| 37 | 188.5 | 13.5 | 333 | 16 | 097H38 | 097H38 pseudomonas |
| 38 | 188 | 13.4 | 299 | 2 | P95448 | P95448 pseudomonas |
| 39 | 188 | 13.4 | 299 | 16 | 09J5M1 | 09J5M1 pseudomonas |
| 40 | 186.5 | 13.3 | 281 | 2 | 09ZGK2 | 09ZGK2 leptospira |
| 41 | 186.5 | 13.3 | 316 | 2 | 093DZ7 | 093DZ7 streptococc |
| 42 | 186 | 13.3 | 277 | 2 | 086296 | 086296 mycobacteri |
| 43 | 186 | 13.3 | 322 | 2 | 09ALX0 | 09ALX0 streptococc |
| 44 | 184 | 13.1 | 411 | 17 | 08VW31 | 08VW31 methanocarc |
| 45 | 183.5 | 13.1 | 281 | 2 | 09S4F8 | 09S4F8 leptospira |

ALIGNMENTS

RESULT 1

088109 ID 088109 PRELIMINARY; PRT; 266 AA.

AC 088109; D8 01-NOV-1998 (TREMUREL. 08, Created)

DT 01-NOV-1998 (TREMUREL. 08, Last sequence update)

DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)

DE GSD Protein.

GN GSD OR GTFD.

OS Mycobacterium avium subsp. silvaticum,

OS Mycobacterium paratuberculosis, and

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=44282, 1770, 1764;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=M.avium subsp. silvaticum, and M.paratuberculosis;

RA Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,

RT "A low G+C content element in Mycobacterium avium subsp.

RT paratuberculosis and M. avium subsp. silvaticum with homologous genes

RT in M. tuberculosis";

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RT [2]

RP SEQUENCE FROM N.A.

RC SPECIES=M.avium; STRAIN=2151;

RA Eckstein T.M., Lambert M.L., Brennan P.J., Bellisle J.T., Inamine J.M.;

RT "Identification of a gene cluster involved in glycopeptidolipid

RT biosynthesis and of a gene cluster encoding daunorubicin resistance in

RT two strains of Mycobacterium avium serovar 2.";

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RT [3]

RP SEQUENCE FROM N.A.

RC SPECIES=M.avium; STRAIN=724;

RA Eckstein T.M., Brennan P.J., Inamine J.M., Bellisle J.T.;

RT "Identification of gene cluster involved in glycopeptidolipid

RT biosynthesis and of a gene cluster encoding daunorubicin resistance in

RT two strains of Mycobacterium avium serovar 2.";

RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ223832; CAAL1574.1; -
 DR EMBL: AJ223833; CAAL1578.1; -
 DR EMBL: AF143772; AAD4222.1; -
 DR EMBL: AF152999; AAD20376.1; -
 DR InterPro: IPR001173; Glycosyltransferase_2.
 DR Pfam: PF00535; Glycosyltransferase_2; 1.
 DR Transferase.
 KW SEQUENCE 266 AA; 30195 MW; F88A2754683F5A8B CRC64;

Query Match 100.0%; Score 1401; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1,2e-114;
 Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTAPESTIIPFNAAVTLQACLSIVGQTYREVEVYLVDSGSDRTLDIANSFRELGS 60
 DB 1 MTAPESTIIPFNAAVTLQACLSIVGQTYREVEVYLVDSGSDRTLDIANSFRELGS 60
 QY 61 RLTVHSGPDGPDGYDAMNRGVATGEMVLEFGADDTLYEPTTLAQVAALFGDHAASHLYV 120
 DB 61 RLTVHSGPDGPDGYDAMNRGVATGEMVLEFGADDTLYEPTTLAQVAALFGDHAASHLYV 120
 QY 121 GDVVRSTKSRHAGPFDRLLEFNLCHQSIFYRRELFDGIGIPYMLRYRVADMPENIR 180
 DB 121 GDVVRSTKSRHAGPFDRLLEFNLCHQSIFYRRELFDGIGIPYMLRYRVADMPENIR 180
 QY 181 CPSNPALTRYNDVYISEYNDMTGFSMROGTKEFKRRLPMYFWAGWETCRMLAFKLD 240
 DB 181 CPSNPALTRYNDVYISEYNDMTGFSMROGTKEFKRRLPMYFWAGWETCRMLAFKLD 240
 QY 241 KENRRLALTRILIRKAVSKERSAEP 266
 DB 241 KENRRLALTRILIRKAVSKERSAEP 266

RESULT 2

QY 093025 PRELIMINARY; PRT: 297 AA.
 AC 093025;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Putative glycosyl transferase Wbdl.
 GN Wbdl.
 OS Salmonella enterica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=28901;
 RN [1]
 RP SEQUENCE FROM N.A.
 Wang L., Reeves P.R.;
 "The E. coli O111 and S. enterica O35 gene clusters: Gene clusters
 RT encoding the same colitose containing O antigen are highly
 RT conserved."
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF285969; AAK83018.1; -
 DR InterPro: IPR001173; Glycosyltransferase_2.
 DR Pfam: PF00535; Glycosyltransferase_2; 1.
 DR Transferase.
 KW SEQUENCE 297 AA; 34136 MW; AAGC9B82D322AB4B CRC64;

Query Match 21.3%; Score 298.5; DB 2; Length 297;
 Best Local Similarity 30.6%; Pred. No. 4,2e-18;
 Matches 72; Conservative 47; Mismatches 81; Indels 35; Gaps 7;

QY 4 PVFSIIIPFNAAVTLQACLSIVGQTYREVEVYLVDSGSDRTLDIANSFRELGSRLV 63
 DB 18 PLISIIIAIFNSELVIANLQSVISQYKNIETIIMDGSSYDKTLDIANSFKKE---RIK 74
 QY 64 VHSQPDGPDGYDAMNRGVATGEMVLEFGADDTLYEPTTLAQVAALFGDHAASHLYV 123
 DB 75 IYSEKDRGIYDAMNRKAVDLSGDMIAFISGSDPYVYHTDAITSYVKGALISNGAPVYGG-- 132

QY 124 VMSTKSRHAGP-----FDLRLLEFNLCHQSIFYRRELFDGIGIPYN 166
 DB 133 -----RTAHGCPMKKEISGSGSEWYNLKGKFNRYKNLPLIPMSATISRDFFD-ORFD 186
 QY 167 LRYRVADMPENIRCF-----SNPALTRYNDVYISEYNDMTGFSMROGTDE 214
 DB 187 IKRKIVADMPFLRCFIKRKENSPIYITDITPVYRMGYGVSTDISQKITTE 241

RESULT 3

QY 095520 PRELIMINARY; PRT: 297 AA.
 AC 095520;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Putative glycosyl transferase.
 GN Wbdl.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-STOKE W;
 RX MEDLINE=96060831; Pubmed-7590310;
 RA Bastin D.A., Reeves P.R.;
 RT "Sequence and analysis of the O antigen gene (rfb) cluster of
 RT Escherichia coli O111."
 RL Gene 164:17-23(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-STOKE W;
 RX MEDLINE=98449835; Pubmed-9774562;
 RA Wang L., Cud H., Qu W., Reeves P.R.;
 RT "Sequencing of Escherichia coli O111 O-antigen gene cluster and
 RT identification of O111-specific genes."
 RL J. Clin. Microbiol. 36:3182-3187(1998).
 DR EMBL: AF078736; AAD46731.1; -
 DR InterPro: IPR001173; Glycosyltransferase_2.
 DR Pfam: PF00535; Glycosyltransferase_2; 1.
 KW Transferase.
 QY SEQUENCE 297 AA; 33740 MW; AF1B86797C2B37CB CRC64;

Query Match 21.3%; Score 298; DB 2; Length 297;
 Best Local Similarity 27.7%; Pred. No. 4,7e-18;
 Matches 78; Conservative 52; Mismatches 86; Indels 66; Gaps 8;

QY 1 MTAPESTIIPFNAAVTLQACLSIVGQTYREVEVYLVDSGSDRTLDIANSFRELGS 60
 DB 15 LDAPLVIIITATVSELDIACLOSQSYKNIETIIMDGSSDKTLDIANSFKFD--- 71
 QY 61 RLTVHSGPDGPDGYDAMNRGVATGEMVLEFGADDTLYEPTTLAQVAALFGDHAASHLYV 120
 DB 72 RIKVISEKDRGIYDAMNRKADSLIGDVAFLIGSDPYVYHTDAITSIMKGVANSGAPVY 131
 QY 121 GDVVRSTKSRHAGP-----FDLRLLEFNLCHQSIFYRREL 158
 DB 132 G-----RTAHGCPDNISGSGSEWYNLGFKNYKMLPLIPMSATISRDFFRNR 184
 QY 159 FDGIGIPYMLRYRVADMPENIRCF-----SNPALTRYNDVYISEYNDMTGFSMROGT 211
 DB 185 FD-----IKRKIVADMPFLRCFIKRKENSPIYITDITPVYRMGYGVSTDISQKITTE 238
 QY 212 DKEFRKRLPMYFWAGWETCRMLAFKDKENRRLALTRILI 253
 DB 239 TLE-----SFYVRKKNNSCLNIQILI 259

RESULT 4
 QY 08VYXO PRELIMINARY; PRT: 265 AA.
 AC 08VYXO;
 DB 08VYXO;

```

DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 31.4 kDa protein.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
CX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOND UTRECHT IV;
RX MEDLINE=21369626; PubMed=11476985;
RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
RT "Genetic differences among the LPS biosynthetic loci of serovars of
RT Leptospira interrogans and Leptospira borgpetersenii.";
RL EMBL; AF316559; AAL49452.1;
DR Interpro: IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Hypothetical protein.
DR SEQUENCE 265 AA; 31424 MW; F85D0153E0197DBB CRC64;
2
Query Match 19.7%; Score 276; DB 2; Length 265;
Best Local Similarity 32.3%; Pred. No.3.4e-16;
Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6
QY 4 PVFSIIPTFNAAVTLQACLSIGVGYREVEVVLVDGSTD-----RTIDIANSEFP 56
DB 6 PRISIIITNLNLEGLRKTLESVKQOTYTNFELIYVDGSDGSGSEFYLSKNDLIRKFI 65
QY 57 ELGSLVYHSGCDGDPDYDANRGVANGEWLFLGADDTLYEPPTLQAAVAFGLDHAAS 116
DB 66 E-----KDKKIIYANQNGISLSKGEYLVFLNAGPTLLQKNVLEISKFLDQD-- 113
QY 117 HLYVDGVYWRSTKSRHAGFEDRLFF-----ETNLCHOSIFRRELFDSIGPNLRYRW 172
DB 114 DLYVNLILDSKDNQIIEKRTIPRLNLYFWYSIKSLCHQAVFLRKSLFDLYGYINEYLPA 173
QY 173 ADWDFENIRCFSPNALITRYMDVIVSEYNDMTGFSM----RQGTDEKFRRLPMYF 223
DB 174 ADFEFHFRWFEKNIKIKRHASVFTVLY-DFNVSVSQPKRKRIAEYQIKKKKYF 227
RESULT 5
O9AEE4 PRELIMINARY; PRT; 265 AA.
O9AEE4
AC 09AEE4;
T 01-JUN-2001 (TREMBLrel. 17, Created)
T 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 31.4 kDa protein.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
CX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LA5;
RX MEDLINE=9023210;
RA Michelson M., Bulach D.M., Vith T., Rajakumar K., Faine S., Adler B.;
RT "Identification and characterization of the dTDP-ribose biosynthesis
RT and transfer genes of the lipopolysaccharide-related rfb locus in
RT Leptospira interrogans serovar Copenhagen1.";
RL J. Bacteriol. 179:1262-1267(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LA5;
RX de la Pena-Moctezuma A., Bulach D.M., Adler B.;
RT "Genetic differences among the LPS biosynthetic loci of serovars of
RT Leptospira interrogans and Leptospira borgpetersenii.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AKIYAMI A;
RX MEDLINE=21369626; PubMed=11476985;

```

```

RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;  

RT "Genetic differences among the LPS biosynthetic loci of serovars of  

RT Leptospira interrogans and Leptospira borgpetersenii.";  

RL FEBS Immunol. Med. Microbiol. 31:73-81(2001).  

DR EMBL; U61226; AAK19913.1; -  

DR EMBL; AF316557; AAL49448.1; -  

DR InterPro; IPR001173; Glycos_transf_2.  

DR Pfam; PF00535; Glycos_transf_2; 1.  

KW Hypothetical protein.  

SQ SEQUENCE 265 AA; 31396 MW; F7CE1E6F781AF59D CRC64;  

  

Query Match 19.6%; Score 275; DB 2; Length 265;  

Best Local Similarity 32.3%; Pred. No. 4.2e-16;  

Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6;  

  

QY 4 PVESITLTFENAAVTLQACISIGVQTFREVVVAVDGSST-----RTLDIANSFRP 56  

DB 6 PRISITITNLNLEGRKLTLESVKQTYTNFELIYVDDGSDTSGSEFYKLSNLDLKRTIS 65  

QY 57 ELGSLRVHSGPDDPYDAMNRGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAS 116  

DB 66 E-----KDKGIYNAQNGKISLSKGEYLVFLNAGDTLLQKNILILEISKFLDQV-- 113  

QY 117 HLIVGDVYMRSTKSKNHAGFPDRLLF-----ETNLCHOSIFRRRLDFGIGIPYNIARVYW 172  

DB 114 DLVYGNILIDSQNGIIEKKYDPRLNIYFWYSIKSLCHQAVFIRKSLFPLDGYGYNEEYLFA 173  

QY 173 ADMVDNICEFSNPALITRYMDVYISEYNDMGFSM----RQTDKEFRRLPMWF 223  

DB 174 ADEEFPHRPEWKNKIKIRHASVFTLV-DPNSVSAQPKNRKRIAEYOKIRKKYF 227  

  

RESULT 6  

.08VYX7 PRELIMINARY; PRT; 265 AA.  

AC 08VYX7;  

DT 01-MAR-2002 (TREMBLrel. 20, Created)  

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  

DE Hypothetical 31.4 kDa protein.  

OS Leptospira interrogans.  

OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.  

OX NCBI_TaxID=173;  

RN [1]  

RP SEQUENCE FROM N.A.  

RC STRAIN=1170.  

RX MEDLINE=21369626; PubMed=11476985;  

RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;  

RT "Genetic differences among the LPS biosynthetic loci of serovars of  

RT Leptospira interrogans and Leptospira borgpetersenii.";  

RL FEBS Immunol. Med. Microbiol. 31:73-81(2001).  

DR EMBL; AF316550; AAL49428.1; -  

DR InterPro; IPR001173; Glycos_transf_2.  

DR Pfam; PF00535; Glycos_transf_2; 1.  

KW Hypothetical protein.  

SQ SEQUENCE 265 AA; 31438 MW; DACEL187B481AF700 CRC64;  

  

Query Match 19.6%; Score 275; DB 2; Length 265;  

Best Local Similarity 32.3%; Pred. No. 4.2e-16;  

Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6;  

  

QY 4 PVESITLTFENAAVTLQACISIGVQTFREVVVAVDGSST-----RTLDIANSFRP 56  

DB 6 PRISITITNLNLEGRKLTLESVKQTYTNFELIYVDDGSDTSGSEFYKLSNLDLKRTIS 65  

QY 57 ELGSLRVHSGPDDPYDAMNRGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAS 116  

DB 66 E-----KDKGIYNAQNGKISLSKGEYLVFLNAGDTLLQKNILILEISKFLDQV-- 113  

QY 117 HLIVGDVYMRSTKSKNHAGFPDRLLF-----ETNLCHOSIFRRRLDFGIGIPYNIARVYW 172  

DB 114 DLVYGNILIDSQNGIIEKKYDPRLNIYFWYSIKSLCHQAVFIRKSLFPLDGYGYNEEYLFA 173

```

QY 173 ADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM-----RQGTDKFERKRLPMYF 223
 DB 174 ADEFFHFRFWMFNKNIKIKHSAVFVTLV-DPFGVSAQPKNRKRRIAEYOKIKKKYF 227

RESULT 7

OBVTW5 PRELIMINARY: PRT; 265 AA.

AC 08VWTW5; 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 31.4 kDa protein.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RN NCBI_TaxID=173;
 RP SEQUENCE FROM N.A.
 RC STRAIN=SMITH.
 RX MEDLINE=21369626; PubMed=11476985;
 RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
 "Genetic differences among the LPS biosynthetic loci of serovars of
 Leptospira interrogans and Leptospira borgpetersenii.",
 FEMS Immunol. Med. Microbiol. 31:73-81(2001).
 DR EMBL: AF316564; AAL49454.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Hypothetical protein.

SEQUENCE 265 AA: 31395 MW; F078BE6F7A3A559D CRC64;

Query Match 19.6%; Score 275; DB 2; Length 265;
 Best Local Similarity 32.3%; Pred. No. 4.2e-16;
 Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6;

OY 4 PVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSTD-----RTLDIANSFRP 56
 DB 6 PFIISITITNLANLEGLKRTLESVKSQTYNFEIIVDGGSTGSGSFYKLSNLDIKKFIS 65
 OY 57 ELGSRIVHSGPDDGPDYDAMNRGVATGEWVLFAGADTLYEPTTLAQVAAFGLDHAAS 116
 DB 66 E-----KDKGIYNAQNGKILLSKGEYLVFLNAGDTLLQKNLLLEISKFLDDV-- 113
 OY 117 HLYVGDVYMRSTKSRAHGFDDLLF---ETNLCHQSIFTRRELFDGIGPYNLRYRW 172
 DB 114 DLYVGNILDSKDNGIIEKRYPDRLNIFYWSIKSLCHQAVFIKSLFDLYGYNEEYLF 173
 OY 173 ADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM-----RQGTDKFERKRLPMYF 223
 DB 174 ADEFFHFRFWMFNKNIKIKHSAVFVTLV-DPFGVSAQPKNRKRRIAEYOKIKKKYF 227

SULT 8

OBVTW9 PRELIMINARY: PRT; 265 AA.

AC 08VWTW9; 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 31.4 kDa protein.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RN NCBI_TaxID=173;
 RP SEQUENCE FROM N.A.
 RC STRAIN=LA1;
 RX MEDLINE=21369626; PubMed=11476985;
 RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
 "Genetic differences among the LPS biosynthetic loci of serovars of
 Leptospira interrogans and Leptospira borgpetersenii.",
 FEMS Immunol. Med. Microbiol. 31:73-81(2001).
 DR EMBL: AF316560; AAL49454.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.

KW Hypothetical protein.
 SQ SEQUENCE 265 AA: 31410 MW; DD9E916D09B77A99 CRC64;

Query Match 19.6%; Score 274; DB 2; Length 265;
 Best Local Similarity 32.9%; Pred. No. 5.1e-16;
 Matches 79; Conservative 40; Mismatches 91; Indels 30; Gaps 7;

OY 1 MTA--PVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSTD-----RTLDIA 51
 DB 1 MTAKEPISITITNLANLEGLKRTLESVKSQTYNFEIIVDGGSTGSGSFYKLSNLDI 60
 OY 52 NSRPPELGSRIVHSGPDDGPDYDAMNRGVATGEWVLFAGADTLYEPTTLAQVAAFGL 111
 DB 61 KRFISE-----KDKGIYNAQNGKILLSKGEYLVFLNAGDTLLQKNLLLEISKFLD 110
 OY 112 DHASHILVYDGVYMRSTKSRAHGFDDLLF---ETNLCHQSIFTRRELFDGIGPYNL 167
 DB 111 ODV--DLYVGNILDSKDNGIIEKRYPDRLNIFYWSIKSLCHQAVFIKSLFDLYGYNE 168
 OY 168 RYRWADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM-----RQGTDKFERKRLPMYF 223
 DB 169 EYLFAADFEFFHFRFWMFNKNIKIKHSAVFVTLV-DPFGVSAQPKNRKRRIAEYOKIKKKYF 227

RESULT 9

OBVT41 PRELIMINARY: PRT; 265 AA.

AC 08VLT41; 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 31.4 kDa protein.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RN NCBI_TaxID=173;
 RP SEQUENCE FROM N.A.
 RC STRAIN=SALINEM, BALLICO, SMART, AND NAAM;
 RX MEDLINE=21369626; PubMed=11476985;
 RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
 "Genetic differences among the LPS biosynthetic loci of serovars of
 Leptospira interrogans and Leptospira borgpetersenii.",
 FEMS Immunol. Med. Microbiol. 31:73-81(2001).
 DR EMBL: AF316556; AAL49446.1; -
 DR EMBL: AF316558; AAL49450.1; -
 DR EMBL: AF316562; AAL49458.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Hypothetical protein.

SEQUENCE 265 AA: 31422 MW; E5539162FA877A9D CRC64;

Query Match 19.3%; Score 271; DB 2; Length 265;
 Best Local Similarity 32.3%; Pred. No. 9.3e-16;
 Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6;

OY 4 PVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSTD-----RTLDIANSFRP 56
 DB 6 PFIISITITNLANLEGLKRTLESVKSQTYNFEIIVDGGSTGSGSFYKLSNLDIKKFIS 65
 OY 57 ELGSRIVHSGPDDGPDYDAMNRGVATGEWVLFAGADTLYEPTTLAQVAAFGLDHAAS 116
 DB 66 E-----KDKGIYNAQNGKILLSKGEYLVFLNAGDTLLQKNLLLEISKFLDDV-- 113
 OY 117 HLYVGDVYMRSTKSRAHGFDDLLF---ETNLCHQSIFTRRELFDGIGPYNLRYRW 172
 DB 114 DLYVGNILDSKDNGIIEKRYPDRLNIFYWSIKSLCHQAVFIKSLFDLYGYNEEYLF 173
 OY 173 ADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM-----RQGTDKFERKRLPMYF 223
 DB 174 ADEFFHFRFWMFNKNIKIKHSAVFVTLV-DPFGVSAQPKNRKRRIAEYOKIKKKYF 227

us-09-705-911-24.rsp

```

Best Local Similarity 31.8%; Pred. NO. 1.9e-15;
Matches 83; Conservative 42; Mismatches 93; Indels 43; Gaps 10;

QY 4 PVFSIIIPFMAAVTLQACISGIVGOTYREVYVVLVDGSDT-----RTLDIANSFRP 56
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6 KRISITITNLNLEBGLRTLESKQOTYTNFELIYVDGSDTGSSEFYELKSNLDLTKRFTS 65
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 57 ELGSRLLVHSGPDDPYDAMNRGVAVATGEWVLFAGADLTLEPTTLQOVAATLGDHAAS 116
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 66 E-----KDKRIYVNAQNGKGIILSKGEYLVPLFAGNALLQKRVLSISKFTLQDNI-- 113
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 117 HLIVGVDVVRSTKSRHACPEDI--DRLEF---ETNLCHOSIYYRRELFDGIGYNTLRY 169
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 DLVYGDILIT--DSKDGHIIERKYPEDRLVYFWSYIKSLCHQATFTLRKNIADLYGYNBEY 170
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 RYVAMQDENINRCFSPALITTRYNDQVVISRYNDMTGFSM-----ROGTDKEFFRRLLMYFVY 225
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 171 LPAADFEFFHRHWEKNKNIKIKHAPVILYI--DPNGVSAQPKRKRKRIAEYRIRKRRKPI 229
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 226 AGWETCRRLMLAFLDKENRRL 246
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 -W-----AYVNKLNSYL 241
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

| | SEQUENCE | 248 AA; | 28560 MW; | C01953610C8DEFAD | CRC64; |
|----|---|--------------|-----------|------------------|--------|
| AD | 09EXY4 | PRELIMINARY; | PRT; | 248 AA. | |
| AC | 09EXY4 | | | | |
| DT | 01-MAR-2001 (TREMBLrel. 16, Created) | | | | |
| DT | 01-MAR-2001 (TREMBLrel. 16, last sequence update) | | | | |
| DT | 01-DEC-2001 (TREMBLrel. 19, last annotation update) | | | | |
| DE | putative glycosyl transferase. | | | | |
| OS | Escherichia coli. | | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | | | | |
| OC | Escherichia | | | | |
| OX | NCBI_TaxID=562; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN-M1182; | | | | |
| RC | MEDLINE=21135136; PubMed=11238967; | | | | |
| RA | Jensen S.O., Reeves P.R.; | | | | |
| RT | "Molecular evolution of the GDP-mannose pathway genes (manB and manC) | | | | |
| RT | in <i>Salmonella enterica</i> ." | | | | |
| RL | Microbiology 147:599-610 (2001). | | | | |
| DR | EMBL; AY012189; AAG41752.1; - | | | | |
| DR | InterPro; IPR001173; GLYcos transf_2. | | | | |
| DR | Pfam; PF00535; GLYcos transf_2.1. | | | | |
| KW | transferase | | | | |

[illegible]

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:21:00 ; Search time 11 seconds

(without alignments)
1002.973 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401
Sequence: 1 MTAFVFSIIPTFNAAVTLQ.....ALRRLIRKAVKSKERSAEP 266

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 788 | 56.2 | 256 | 1 Y757_MYCTU | O50459 mycobacteri |
| 2 | 202.5 | 14.5 | 262 | 1 Y714_MYCTU | P71793 mycobacteri |
| 3 | 190.5 | 13.6 | 322 | 1 YS86_AMASP | P22639 anabaena sp |
| 4 | 172.5 | 12.3 | 344 | 1 Y1BD_ECOTL | P11390 escherichia |
| 5 | 168 | 12.0 | 250 | 1 Y868_HAEIN | O57022 haemophilus |
| 6 | 165 | 11.8 | 323 | 1 YF78_HAEIN | O57287 haemophilus |
| 7 | 147.5 | 10.5 | 248 | 1 WCAE_ECOLI | P71339 escherichia |
| 8 | 146.5 | 10.5 | 1275 | 1 REBC_MYXXA | O50864 myxococcus |
| 9 | 141 | 10.1 | 301 | 1 AMSB_ERWAM | O46632 erwinia amy |
| 10 | 139 | 9.9 | 294 | 1 YG96_HAEIN | O48214 haemophilus |
| 11 | 138.5 | 9.9 | 268 | 1 YMOF_BACSU | P39614 bacillus su |
| 12 | 136 | 9.7 | 441 | 1 YCDO_ECOTL | P75905 escherichia |
| 13 | 134 | 9.6 | 342 | 1 EXOU_RHIME | P33700 rhizobium m |
| 14 | 134 | 9.6 | 348 | 1 EXOO_RHIME | P33697 rhizobium m |
| 15 | 131.5 | 9.4 | 258 | 1 MAAE_RLEPN | O9xc90 klebsiella |
| 16 | 129.5 | 9.2 | 299 | 1 YAGI_RHISN | P55465 rhizobium s |
| 17 | 127.5 | 9.1 | 299 | 1 Y025_MYCPN | P43086 mycoplasma |
| 18 | 127.5 | 9.1 | 900 | 1 GGAB_BACSU | P46918 bacillus su |
| 19 | 124.5 | 8.9 | 256 | 1 SP5A_BACSU | P39621 bacillus su |
| 20 | 124.5 | 8.9 | 319 | 1 YF18_MYCTU | O50590 mycobacteri |
| 21 | 124.5 | 8.9 | 446 | 1 GGAA_BACSU | P46917 bacillus su |
| 22 | 124 | 8.9 | 298 | 1 Y025_MYCCE | P47271 mycoplasma |
| 23 | 122.5 | 8.7 | 299 | 1 Y060_MYCPN | P75042 mycoplasma |
| 24 | 119.5 | 8.5 | 426 | 1 CGED_BACSU | P42092 bacillus su |
| 25 | 118.5 | 8.5 | 297 | 1 Y060_MYCCE | P47306 mycoplasma |
| 26 | 113 | 8.1 | 403 | 1 YCOA_SYNP7 | P43460 synechococc |
| 27 | 109.5 | 7.8 | 333 | 1 REBV_SALTY | P24401 salmonella |
| 28 | 108 | 7.7 | 290 | 1 YAS7_MERTH | O38457 methanococc |
| 29 | 105.5 | 7.5 | 241 | 1 DPM1_DROME | O97417 desoriphila |
| 30 | 104 | 7.4 | 279 | 1 WCAE_ECOLI | P77414 escherichia |
| 31 | 102.5 | 7.3 | 267 | 1 YG95_HAEIN | O48215 haemophilus |
| 32 | 102 | 7.3 | 346 | 1 YF20_MYCTU | O50587 mycobacteri |
| 33 | 101 | 7.2 | 340 | 1 REB4_ECOTL | P13961 escherichia |

| | | | | | |
|----|------|-----|-----|--------------|--------------------|
| 34 | 100 | 7.1 | 343 | 1 REP7_ECOLI | O52221 escherichia |
| 35 | 100 | 7.1 | 343 | 1 REP2_ECOLI | P18023 escherichia |
| 36 | 99.5 | 7.1 | 314 | 1 REBN_SALTY | P26403 salmonella |
| 37 | 99.5 | 7.1 | 426 | 1 NODC_RHIME | P03431 rhizobium m |
| 38 | 94.5 | 6.7 | 398 | 1 YAIR_ECOTL | O47536 escherichia |
| 39 | 94 | 6.7 | 395 | 1 NODC_AZOCA | O07765 azorhizobiu |
| 40 | 94 | 6.7 | 419 | 1 HASA_STRPY | O34865 streptococc |
| 41 | 93.5 | 6.7 | 236 | 1 DPM1_SCHPO | O14466 schizosacch |
| 42 | 92.5 | 6.6 | 260 | 1 DPM1_HUMAN | O60762 homo sapien |
| 43 | 92 | 6.6 | 266 | 1 AMSB_ERWAM | O46635 erwinia amy |
| 44 | 92 | 6.6 | 424 | 1 NODC_RHITO | P17862 rhizobium l |
| 45 | 90.5 | 6.5 | 254 | 1 Y653_HAEIN | P44023 haemophilus |

ALIGNMENTS

RESULT 1
ID Y757_MYCTU STANDARD: PRT; 256 AA.
AC O50459;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase RV2957 (EC 2.-.-.-).
GN RV2957 OR MT3031 OR MTCY349.31C OR U0002KC.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae (Class); Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID-1773;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-H37RV;
RX MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT *Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.*
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN-CDC 1551 / Oshkosh;
RC Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Holt D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT *Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
DR EMBL: U00024; AAA50938.1; AUT_INIT.
DR EMBL: 283018; CAB05419.1; AUT_INIT.
DR EMBL: AE007125; AAK47357.1; -.

DR TIGR; MT3031; -.
DR TUBerculin: Rv2957; -.
DR InterPro: IPR001173; glycos_transf.2.
DR Pfam: PF00535; glycos_transf.2; 1.
KM Hypothetical protein; glycosyltransferase
KM Complete Proteome. 236 AA; 29012 MW; 90C6EC628C59CA57 CRC64;
50 SEQUENCE

| | | | | |
|-----------------------|------------------|----------------|------------|-------------|
| Query Match | 56.2%; | Score 788; | DB 1; | length 256; |
| Best Local Similarity | 60.9%; | Pred. No. | 3.3e-65; | |
| Matches 156; | Conservative 24; | Mismatches 54; | Indels 22; | Gaps 2 |

```

QY      1  MTAVEFSIIITTFMAAATVLAOCAGSIYGCITYRREVEVAVDVGSGSDRLDIANSFPRELS  60
Db      1  MAAMFSTIIPTLVAAVLAIVPRCLDSIAKQCISGPELVAVDVGSGTDELDIANIFAPNLGE  60

QY      61  RLVAHSGFDGDPYDAMNRGVAVATGEVWVLELGADDTLYEPTTLAQAFAFGDHAASHLYV  120
Db      61  RLIIHRPTDQCVYDAMNRGVDAIAGTWLLEFGADDSIYEADPTLARAFAVIGEHEPSPDLYV  120

QY      121  GDVYMRSTKSHAAPEFLDLRLFEPTNLCHOSTIRYRELEPFGIGPYNLRVYVADMPENIR  180
Db      121  GDVIMRSTNFMFGAFLDRLRLFRKATICHQATLIRKGLFSTIGYINLIRYVLDMPDNIR  180

QY      181  CFSNPALITRYMDVIVSEYNDMTGFSNRQGTDEFRKRLRMFYFVAVAGWETCRMLAFLKD  240
Db      181  CFSNPALVTRYMHVAVASYNDEGLS-NTLVDEKEFLKRLPM-----  220

QY      241  KENRRIALRTLLRYK  256
Db      221  --STRLGRIYVILVR  234

```

| RESULT 2 | FE14_MYCTU | STANDARD; | PRT; | 262 AA. |
|----------|--|-----------|------|---------|
| ID | FE14_MYCTU | | | |
| AC | P11793; | | | |
| DT | 01-NOV-1997 (Rel. 35, Created) | | | |
| DT | 15-NOV-1997 (Rel. 35, Last sequence update) | | | |
| DT | 01-JUN-2002 (Rel. 41, Last annotation update) | | | |
| DE | Putative glycosyl transferase R15154c (EC 2.-.-.-). | | | |
| GN | R15154C OR MT1564 OR MTC1277.36C. | | | |
| OS | Mycobacterium tuberculosis. | | | |
| OC | Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; | | | |
| OC | Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium | | | |
| OX | NCBI_TaxID:1773; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=H37RV; | | | |
| RC | MEDLINE=98295987; PubMed=9634230; | | | |
| RY | | | | |
| A | Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D. | | | |
| A | Gordon S.V., Eiglmeyer K., Goss S., Barry C.E. III, Tekala F., | | | |
| A | Bacchok R., Basham D., Brown D., Chillingworth T., Connor R., | | | |
| A | Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S., | | | |
| A | Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., | | | |
| A | Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., | | | |
| A | Rutter S., Seeger K., Skelton S., Squares S., Barrett B.G., | | | |
| A | Sulston J.E., Taylor K., Whitehead S., Barrell B.G.; | | | |
| RT | *Deciphering the biology of Mycobacterium tuberculosis from the | | | |
| RT | complete genome sequence. *; | | | |
| RL | Nature 393:537-544(1998). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=CDC 1551 / Oshkosh; | | | |
| RC | Eletschmann R.D., Alland K., Eisen J.A., Carpenter L., White O., | | | |
| RA | Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E., | | | |
| RA | Kolonyak J.F., Nelson W.C., Umayam L.A., Emtolaeva M.D., Salzberg S.L. | | | |
| RA | Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., | | | |
| RA | Bisbal W.; | | | |
| RT | *Whole genome comparison of Mycobacterium tuberculosis clinical and | | | |
| RT | laboratory strains. *; | | | |
| RL | Submitted (APR-2001) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2. | | | |

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC
DR EMBL; Z79701; CAB02028.1; -.
DR EMBL; AE007023; AAK45831.1; -.
DR TIGR; MT1564; -.
DR Tuberculist; Rv151Ac; -.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
RV Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 262 AA; 28965 MW; 6B229BF8D31923E75 CMC64;

```

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 14.5% | Score 202.5; | DB 1; | Length 262; |
| Best Local Similarity | 32.2% | Pred. No. 2,2e-11; | | |
| Matches 68: | Conservative 36; | Mismatches 90; | Indels 17; | Gaps 9 |

```

QY 2 TAPFESIITPEFNAAVTLQALCLGIVGCTY-EEVWLVNDSGSDPRLDIANSRPEIGS 60
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 3 SAPFVSVTLTISFNDLDGIQRTVKSQVRAORTXGRILEHYIDGSGDDVAVYISGCEPFA- 61
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 61 RLTVHSGPDDPGPYDAMNKGVSATGEWVLEFGADDTLYEPTTLAQ-VAAFGSDRAASHLV 119
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 62 --WYMSBPDGGRYDAMNGINAHSGDILMTLHSAHDRSGPDVAAQAEALSGKGPSEL- 118
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 120 YG----DYVMASSTKSRHAGPFDLRLLEFTNLC-HQSIFYRRELFDGIGPYLRRVADM 175
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 119 WGFQNDRLVGLDRAKGRPIPELRLREFLAGKQVPPHQASFEFGSSLVAKIGGYDLDEGIAAQ 178
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 176 DENIRCFSPALITRYMDV--VISEYNDMG 204
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 179 EFLIRA----ALVCEPVYIRCVLCEE--DTTG 204
    1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1

```

| ID | YS86_ANASP | STANDARD; | PRT; | 322 AA. |
|----|---|-----------|------|---------|
| AC | P22639. | | | |
| DT | 01-AUG-1991 (Rel. 19, Created) | | | |
| DT | 15-JUN-2002 (Rel. 41, Last sequence update) | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | |
| DE | Putative glycosyl transferase alr2836 (EC 2.-.-.-). | | | |
| GN | ALR2836. | | | |
| OS | Anabaena sp. (strain PCC 7120). | | | |
| OC | Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc. | | | |
| OX | NCBI_TaxID:103690; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=21595285; PubMed=11759840; | | | |
| RA | Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S., | | | |
| RA | Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T., | | | |
| RA | Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A., | | | |
| RA | Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M., | | | |
| RA | Yasuda M., Tabata S.; | | | |
| RT | "Complete genomic sequence of the filamentous nitrogen-fixing | | | |
| RT | cyanobacterium Anabaena sp. strain PCC 7120."; | | | |
| RL | DNA Res. 8:205-213(2001). | | | |
| RN | [2] | | | |
| RX | SEQUENCE OF 1-131 FROM N.A. | | | |
| RX | MEDLINE=90264305; PubMed=2111805; | | | |
| RA | Holland D., Wolk C.P.; | | | |
| RA | "Identification and characterization of heta, a gene that acts early | | | |
| RA | in the process of morphological differentiation of heterocysts."; | | | |
| RL | J. Bacteriol. 172:3131-3137(1990). | | | |
| CC | -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2. | | | |
| CC | THIS SWISS-PROT entry is copyright It is produced through a collaboration | | | |


```

DB 123 YGQVNIIDGSLKVMPLKPVGEHSIAEYLFQCYG-FIOTSTIVLKREDAEIRFD--- 180
QY 143 FETNICHOSIEYFRELFDGIPYLNRYRWADWDNRCFSPNALITRYMDVY--ISEVN 200
DB 181 -BRYIRHQ-----DYDL---CIRADKLGFKFMVINOPLANYH 213
QY 201 DMTGF-SMOCGTDKER-----RKLEPMYFVWAGETCRRLA-- 236
DB 214 MYTRGSOHKGESVKYSLFLWLDAMKPHLTRRDVYTYKAYKLPLRYKMDG-KSLQASLSFA 272
QY 237 ---FLDKENR-----RL--ALRRL 252
DB 273 RYFFLELNDRNDELKRLANKLRTLR 298

RESULT 10
YQ96_HAEIN STANDARD: PRT: 294 AA.
AC 048214: 005082;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
Putative glycosyl transferase H1696 (EC 2.---).
H1696
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
NX NCBI_TaxID=727;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=42;
RA McLaughlin R., Abu Kwak Y., Young R., Spinoia S., Apicella M.;
RT "Characterization and sequence of the lsg locus from Haemophilus
RL influenzae.";
RN Submitted (JUN-1992) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fline L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gresham C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RN "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
-1 SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M94855; AAA24982.1; -
DR EMBL: U32842; AAC23342.1; -
DR TIGR: H11696; -
DR InterPro: IPR001173; Glycos transf_2.
DR Pfam: PF00535; Glycos_transf_2.1.
KM Hypothetical protein; Transferase; Glycosyltransferase;
KM Complete proteome.
FT CONFLICT 38 C -> Y (IN REF. 1).
FT CONFLICT 48 S -> R (IN REF. 1).
FT CONFLICT 70 V -> I (IN REF. 1).
FT CONFLICT 74 T -> A (IN REF. 1).
FT CONFLICT 86 C -> R (IN REF. 1).

```

```

FT CONFLICT 97 97 V -> I (IN REF. 1).
FT CONFLICT 106 106 N -> D (IN REF. 1).
FT CONFLICT 152 152 I -> V (IN REF. 1).
FT CONFLICT 185 185 P -> S (IN REF. 1).
SQ SEQUENCE 294 AA; 33646 MW; 8330F081BFFA18B CRC64;

Query Match 9.98; Score 139; DB 1; Length 294;
Best local similarity 23.08; Pred. No. 1.7e-05;
Matches 44; Conservative 40; Mismatches 93; Indels 14; Gaps 5;

QY 5 VFSTIIITFENAATVLAQCLSGIVGQTYREVAVLVADGSDTRTDIANSPRELGSRLV 64
DB 1 MLSTIVSYNKRKAEVPALESRLTQOTSNSFPIIYDDCSKRYV-VEQSYSEFPV---TVI 56
QY 65 HSGPDDGPYDAMNRGVATGEVWLFGLGADPTLVEPTTLAQAVALFGHAAHSLVY---- 120
DB 57 RNEFNOGAESRNVGARTSKGDMFLFLD-DDDCRPEKCEVLOYIEQNPINIFHPAK 115
QY 121 ----GDVVMRSTKSRHAGPFDDLRLFEETNLCHO-SIFRRELFDGIPYLNRYRWADW 175
DB 116 CEMVNEGFYVYTOPLEPOEISTERILLANKIGCPMPMIAIKKEMFLKIGLSTALRSLEDY 175
QY 176 DFNRCFSNPA 186
DB 176 DFLKLQLQEPS 186

RESULT 11
YWDF_BACSU STANDARD: PRT: 268 AA.
AC P39614;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase ywdf (EC 2.---).
GN YWDF OR IPA-56D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presacan E., Dancu A., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Dancu A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RL kb region from 325 degrees to 333 degrees.";
RN Mol. Microbiol. 10:371-384(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Besseiers P., Bolydin A., Borchert S.,
RA Brouillet S., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.V., Carter N.M.,
RA Denicot F., Devyne K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA Entlin R.D., Errington J., Fabret C., Ferrari E., Fouliger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi E.J., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hachez J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nohack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

```

RA Rieger M., Rivolta C., Rocca E., Rocche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schoeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretto A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weizenegeer T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,
 RT The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.
 RT Nature 390:249-256(1997).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X73124; CAB51612.1; -
 DR EMBL: Z69123; CAB51824.1; -
 DR PIR: S39711; S39711.
 DR Subtilisin; BG10602; ywdf.
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 DR Hypothetical protein; Transferase; Glycosyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 268 AA: 30616 MW: DD6428F7016EC9B3 CRC64;
 Query Match 9.9%; Score 138.5; DB 1; Length 268;
 Best Local Similarity 22.5%; Pred. No. 1.7e-05;
 Matches 61; Conservative 52; Mismatches 115; Indels 43; Gaps 10;
 QY 7 STIIPENAAVTLQACLSIVGQTYREVEVYLV-DGSGTRTIDIANSPREPGLSRLVH 65
 DB 4 STIIVYNNIPALCELLLESISRTQMPYEITIIYNDAG--ESVVPKALVPEL-PIAVIN 59
 QY 66 SGPDGDPYDAMNKGAVATGEVWFLFGADDTLEPTTIAOVAFLGDHAASHLYVGPVMA 125
 DB 60 LEKNSGVAARNAKGVKASGDCIM-LCDDDEFTPGHIEKAK--EIEIADYHSAE1 115
 QY 126 RSTKSHAGPFDRLLEFETNLCHOSI-----FYRRELFGIGPNRYRMAD 174
 DB 116 VSEFEENKGRYPRYSKLFATADYEDMRVSTYVPSGMYRFLHDEIGYDADVHNYMD 175
 QY 175 KDFNICEFN-----PALITRYMDVYISEYNDMTGFSMRGOTDEKFRK---RLPMY 222
 DB 176 WDFYLAADYRVKRVPCASVIYAFSDAGNOSADI-GAKRKQYLDRLSEKHGIGELP-- 232
 QY 223 FWVAGMETCRMLAFLKDKENRRLALTRLI 253
 DB 233 -----TKNFAVLEPEPMKREKSEMY 255
 RESULT 12
 ID YCDO_ECOLI STANDARD: PRT: 441 AA.
 AC P75905;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Hypothetical protein ycd0.
 GN YCDO OR B1022.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RT Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horikuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RT DNA Res. 3:137-153(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO Y.PESTIS HMS LOCUS PROTEIN HMSR AND TO
 CC S.EPIDERMIDIS ICA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AE000204; AAC74107.1; -
 DR EMBL: D90739; BAA35803.1; -
 DR EcoGene: EG13863; ycd0.
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 5 25 POTENTIAL.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 331 351 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 395 415 POTENTIAL.
 SQ SEQUENCE 441 AA: 50765 MW: 04F5A53D72FEBAB CRC64;
 Query Match 9.7%; Score 136; DB 1; Length 441;
 Best Local Similarity 19.9%; Pred. No. 5.4e-05;
 Matches 54; Conservative 41; Mismatches 101; Indels 76; Gaps 8;
 QY 4 PVFSIIPTFNAAVTLQACLSIVGQTYREVEVYLV-DGSGTRTIDIANSPREPGLSRLV 63
 DB 75 PSISIIIPCFNEKNEVETIIAALAQRENEIVAVNGSDKRALIDMAAQPHRLV 134
 QY 64 VHSQPDGDPYDAMNKGAVATGEVWFLFGAD-----DTLEPTT----- 102
 DB 135 ITHAONOGKALALKTGAANAASEYLVICIDGALLDRDAAYIVPMALPRVGAATGNPR 194
 QY 103 -----LAOVAFLGDHAASHLYVGDV-----VMSTKSRHAGPF----- 136
 DB 195 IRTSTLVGKIQVGEYSIIILIKRTQRIYGVTVSVIAFRSALAEGVWSDDMIT 254
 QY 137 -DIDRLLEFETNLCHOSIYR-----RELFGIGPNRYR-----RYWA 173
 DB 255 EDID-ISMKQLQNMOTIYEPBALCMILMPETLGLMKORLMAOGGAEEVFLKNMTRLMR 313
 QY 174 DMDPNICFSNPALITRYMDVYISEYNDMTGE 205
 DB 314 KENFRMW-----PFEFCYLTITWAFCTLVGF 340
 RESULT 13
 ID EXOU_RHIME STANDARD: PRT: 342 AA.
 EXOU_RHIME

```

AC P33700;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinoglycan biosynthesis protein exou (EC 2.2.1.11).
GN EXOU OR RB1071 OR SMC20948.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT succinoglycan by Rhizobium meliloti."
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / S047;
RX MEDLINE=94162682; PubMed=8118055;
RA Becker A., Kleckmann A., Kuester H., Keller M., Arnold W.,
RA Puehler A.;
RT "Analysis of the Rhizobium meliloti genes exou, exov, exow, exot, and
RT exoi involved in exopolysaccharide biosynthesis and nodule invasion:
RT exou and exow probably encode glycosyltransferases."
RL Mol. Plant Microbe Interact. 6:735-744(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC SUCCINOGLYCAN (EPS I). NEEDED FOR THE ADDITION OF THE SIXTH SUGAR
CC (GLUCOSE). CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN
CC THE FIFTH AND SIXTH SUGAR.
CC -1- PATHWAY: Exopolysaccharide biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: L20758; AAA16053.1;
DR EMBL: 222646; CAAB0359.1;
DR EMBL: AL603645; CAC49471.1;
DR PIR: D49348; D49348.
DR InterPro: IPR001173; Glycosyltransf.2.
DR Pfam: PF00535; Glycosyltransf.2.1.
DR Transferrase; Glycosyltransferase; Exopolysaccharide synthesis;
KM Plasmid; Complete proteome.
SQ SEQUENCE 342 AA; 37017 MW; EAF55E0EBDA023BC CRC64;

```

```

Query Match 9.6%; Score 134; DB 1; Length 342;
Best Local Similarity 25.3%; Pred. No. 6e-05;
Matches 66; Conservative 38; Mismatches 125; Indels 32; Gaps 10;

```

```

QY 8 IITPFAVTLQACIGSTVQTYREVVVLVDGSGTDRDLIDANSFRPDLGRLVAVHS 67
   ||| ||| : : : ||| : ||| : : : ||| : ||| : : : ||| : |||
DB 11 IISKNAADTITARAVALALAP-ELAEVYVLDGSGTSDSASVARAADGTGLNVVRF 65

```

```

QY 68 PDDGPDAMRGVATGENTVLEFGADDTLYEPTTLAQVAFLG-DHAASHLYGDVYMR 126
   ||| ||| : : : ||| : ||| : : : ||| : ||| : : : ||| : |||
DB 70 ENGGPAARHAAHAISPLIGVLDADDPFF-PPRLQQLSQDQWDFADNIAIAQA 128
QY 127 STSRHAGGP-DIDRLI-----FETNCHOSI-----FYRRELFDSIG-PYNLRY 169
   ||| ||| : : : ||| : ||| : : : ||| : ||| : : : ||| : |||
DB 129 ATAHGRDRAPPPRLDLVGVFEGNISRGVRGEGTGLKPLRRAPFLDQHGRLVETL 188
QY 170 RVAMDMFNIRCFSPNLTIRYMDVVISSEYNDMTGFSMRQCTDKEF-RKRLPMFVWAGW 228
   ||| ||| : : : ||| : ||| : : : ||| : ||| : : : ||| : |||
DB 189 RLEDYDLVARALANGA---RYKTIHSCGYAAVVRGNSLSGSHRTIDLKRL-----Y 237
QY 229 ETCRRMLATLKDKENRRRLAR 249
   ||| ||| : : : ||| : ||| : : : ||| : ||| : : : ||| : |||
DB 238 EADRAILAGSRLSSDAEAVR 258

```

```

RESULT 14
ID EXOU_RHIME STANDARD; PRT; 348 AA.
AC P33697;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinoglycan biosynthesis protein exou (EC 2.2.1.11).
GN EXOU OR RB1084 OR SMC20959.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT succinoglycan by Rhizobium meliloti."
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / S047;
RX MEDLINE=21396508; PubMed=11481431;
RA Becker A., Kleckmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the Rhizobium meliloti exoMOMP genes
RT involved in exopolysaccharide biosynthesis and mapping of promoters
RT located on the exoHKLAMP fragment."
RL Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / S047;
RX MEDLINE=94067019; PubMed=8246891;
RA Becker A., Kleckmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the Rhizobium meliloti exoMOMP genes
RT involved in exopolysaccharide biosynthesis and mapping of promoters
RT located on the exoHKLAMP fragment."
RL Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC SUCCINOGLYCAN (EPS I). NEEDED FOR THE ADDITION OF THE FIFTH SUGAR
CC (GLUCOSE). CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN
CC THE FOURTH AND FIFTH SUGAR.
CC -1- PATHWAY: Exopolysaccharide biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 208.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L20758; AAA16044.1; ALT_FRAME.

DR EMBL; Z22636; CAA80347.1; -

DR EMBL; AL603645; CAC49484.1; -

DR PIR; C49348; C49348.

DR PIR; S37029; S37029.

DR PIR; S39958; S39958.

DR InterPro; IPR001173; Glycos_transf_2.

DR Pfam; PF00535; Glycos_transf_2; 1.

KW Transferase; Glycosyltransferase; Exopolysaccharide synthesis;

KW Plasmid; Complete proteome.

FT DOMAIN 38 45

SQ SEQUENCE 348 AA; 38131 MW; 770B43782F785579 CRC64;

Query Match 9.6%; Score 134; DB 1; Length 348;

Best Local Similarity 24.0%; Pred. No. 6.1e-05;

Matches 69; Conservative 47; Mismatches 129; Indels 42; Gaps 11;

QY 4 PVEFTIIPFNAAVTLQACLSIVGQTYREVEVVLVDGSGTDRITDIANSFPELGSRV 63

DB 10 PVEFTVVAAYNSADTVIRAIISALQEGVYEVVVVDDCSADATPALVAI-PDPRVRLI 68

QY 64 VHSQDDGPYDAMNKGVAATGCVLFLGADPTLYEPTTLAQVAAFLGDHAAHLVYGV 123

DB 69 A-LDRNRGCGGARNGAGARNGRIAVLSDSDTV-RPDRLRMIIE-RADAGAGQIAVDNL 125

QY 124 VNRSTKSRHAGPF--DDRL-----LFTNLCHOS-----IPYRRLPDGIGP 164

DB 126 DVYSLDGSLSRFSAEIARLPQLLPATIESNVLFRSHNGYMKPIERERLENGQLR 185

QY 165 YNLRYVADMDP-----NIRCFSNPA--LITRYMDVISEYNDMTGFSMGTGKEF 215

DB 186 FDEALRIGEDYILLASALACGRCVAPESAGYIYHIREGSIISVRLDHIIDAMIADEAF 245

QY 216 KRRLEP-----YFVAGMETCRMAFLDKENRRL--ALRIPL 252

DB 246 LRRYALDGLAQRMQHRMRMGFRFARSFLVLEQLKRLSLAGALKTAL 292

RESULT 15

WAAE_KLEPN STANDARD; PRT; 258 AA.

AC Q9XC90;

DT 30-MAY-2000 (Rel. 39, Created)

RT 30-MAY-2000 (Rel. 39, Last sequence update)

RL 15-JUN-2002 (Rel. 41, Last annotation update)

DE Lipopolysaccharide core biosynthesis glycosyl transferase waaE (EC 2.-.-.-).

GN WAAE.

OS Klebsiella pneumoniae.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Klebsiella.

OX NCBI_TaxID=573;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C3;

RX MEDLINE=21264367; PubMed=11371519;

RA Altarriba M., Climent N., Abitia N., Coderech N., Merino S., Izquierdo L.,

RT "Genetic characterization of the Klebsiella pneumoniae waa gene cluster, involved in core lipopolysaccharide biosynthesis.";

RL J. Bacteriol. 183:3564-3573(2001).

CC -I- PATHWAY: Lipopolysaccharide core biosynthesis.

CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2. WAAE/KDTX

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF146532; AAD37772.1; -

DR InterPro; IPR001173; Glycos_transf_2.

DR Pfam; PF00535; Glycos_transf_2; 1.

KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.

SQ SEQUENCE 258 AA; 29268 MW; D745889903067BD9 CRC64;

Query Match 9.4%; Score 131.5; DB 1; Length 258;

Best Local Similarity 25.8%; Pred. No. 7.1e-05;

Matches 47; Conservative 32; Mismatches 46; Indels 57; Gaps 9;

QY 7 SIIPFNAAVTLQACLSIVGQTYREVEVVLVDGSGTDRITDIANSFPELGSRV 66

DB 6 SYVMIAKNADLLPDCLGSV--SMAD-EIIVLDGSGTDMVELAR--RLGAQVYTHT 57

QY 67 GPDDGPYDAMNKGVAATGCVLFLGADPTLYEPTTLAQVAAFLGDHAAHLVYGV 108

DB 58 D-----WRGIGIQRRQADYATGDWLMIDTDERVTPPELRQALIKYLDAPRGATY 101

QY 102 TLAQVAAFLGDHAAHLVYGVVNRSTKSRHAGPFDRLLFTNLCHOSIFRRELEPDG 161

DB 109 SIARNTFLGRFMRHSGWTFDRVRL-----YERARIRYNDNLVHESL-----DS 153

QY 162 IG 163

DB 154 LG 155

Search completed: April 17, 2003, 17:27:24

Job time : 24 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:25:05 ; Search time 47 Seconds
(without alignments)
544.080 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401

Sequence: 1 MTAPEVSIILIPFNAAVTIQ.....ALTRLIRYKAVSKERSAPP 266

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match length | ID | Description |
|------------|-------|--------------------|-------|---------------------------|
| 1 | 788 | 56.2 | 275 2 | B70670 hypothetical prote |
| 2 | 238.5 | 11.0 | 247 2 | AG0376 probable glycosyl |
| 3 | 232.5 | 16.6 | 260 2 | E71975 hypothetical prote |
| 4 | 231.5 | 16.5 | 259 2 | F64532 conserved hypotet |
| 5 | 220 | 15.7 | 324 2 | AB2190 hypothetical prote |
| 6 | 207 | 14.8 | 321 2 | AG2188 hypothetical prote |
| 7 | 206.5 | 14.7 | 316 2 | AE2189 hypothetical prote |
| 8 | 205.5 | 14.7 | 298 2 | B75096 glycosyl transfera |
| 9 | 204.5 | 14.6 | 330 2 | AE2188 glycosyl transfera |
| 10 | 202.5 | 14.5 | 262 2 | E70714 hypothetical prote |
| 11 | 196.5 | 14.0 | 337 2 | AG1920 hypothetical prote |
| 12 | 196.5 | 14.0 | 367 2 | G95948 probable glycosyl |
| 13 | 192 | 13.7 | 336 2 | AG168 glycosyltransfera |
| 14 | 190.5 | 13.6 | 322 2 | AE2160 glycosyltransfera |
| 15 | 190.5 | 13.6 | 323 2 | AD2189 glycosyltransfera |
| 16 | 190 | 13.6 | 248 2 | C90984 hypothetical prote |
| 17 | 190 | 13.6 | 248 2 | F85829 probable glycosyl |
| 18 | 189.5 | 13.5 | 333 2 | AE2026 glycosyl transfera |
| 19 | 188.5 | 13.5 | 333 2 | B97168 glycosyltransfera |
| 20 | 188 | 13.4 | 299 2 | B83557 probable glycosyl |
| 21 | 182 | 13.0 | 344 2 | AC0974 probable glycosyl |
| 22 | 178 | 12.7 | 356 2 | S74766 hypothetical prote |
| 23 | 177.5 | 12.7 | 343 2 | AT2091 glycosyltransfera |
| 24 | 175 | 12.5 | 318 2 | AG2189 glycosyltransfera |
| 25 | 174.5 | 12.5 | 333 2 | G86651 sugar transferase |
| 26 | 174.5 | 12.5 | 623 2 | ARI209 B. subtilis minor |
| 27 | 174.5 | 12.5 | 996 2 | S76194 hypothetical prote |
| 28 | 173.5 | 12.4 | 324 2 | A69290 probable glycosyl |
| 29 | 173.5 | 12.4 | 338 2 | E91190 probable regulator |

| | | | | |
|----|-------|------|--------|---------------------------|
| 30 | 173.5 | 12.4 | 338 2 | F86037 probable regulator |
| 31 | 172.5 | 12.3 | 344 1 | O3ECMH hypothetical 40.5k |
| 32 | 172 | 12.3 | 573 2 | B89789 hypothetical prote |
| 33 | 171 | 12.2 | 344 2 | A70037 capsular polysacch |
| 34 | 170.5 | 12.2 | 301 2 | F95205 glycosyl transfera |
| 35 | 170 | 12.1 | 354 2 | H96021 probable glycosyl |
| 36 | 169.5 | 12.1 | 774 2 | AC1566 hypothetical prote |
| 37 | 168.5 | 12.0 | 294 2 | G71148 probable glycosyl |
| 38 | 168 | 12.0 | 250 2 | A64039 glycosyl transfera |
| 39 | 168 | 12.0 | 318 2 | AH2189 glycosyl prote |
| 40 | 168 | 12.0 | 462 2 | AA1566 galactosamine-cont |
| 41 | 168 | 12.0 | 1013 2 | AE1876 hypothetical prote |
| 42 | 165.5 | 11.8 | 340 2 | T44330 glycosyl transfera |
| 43 | 165 | 11.8 | 303 2 | AB4114 glycosyl transfera |
| 44 | 165 | 11.8 | 323 1 | H64130 glycosyl transfera |
| 45 | 164 | 11.7 | 334 1 | G71153 hypothetical prote |

ALIGNMENTS

RESULT 1

B70670 hypothetical protein Rv2957 - Mycobacterium tuberculosis (strain H37RV)

N:Alternate names: u0002kc protein

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence.revision 17-Jul-1998 #text_change 20-Jun-2000

0Y 7 SIIPFNAVTLQACLSIVGQTYREVEVLVDGGSTDRTLDIANSFRPELGSRLVHS 66

R,Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyama, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata-Nakazaki, S. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

[illegible]

QY 61 ----RLVHSGPDDGPRYDAMNRCGVATGEMVLEFLGADDTLYEPTTLAQVAALFGDHAAS 116
 DB 59 IRYTRLRKNSG---GPI-ARNIGIKKAKGRFTALLDDDEMLPHRLEVYVRFENLGEKEF 114
 QY 117 HLVDVYMRSTKSR-----HAGPFDDLRLFPETNLCIOSIFRYRRLFPGIGPYNR 168
 DB 115 GVVYGGFYVSDGRIILGRRLPKHNGDI-YSHLKENFISGPTLLIRRCFKAAGLFDPR 173
 QY 169 YRWADMDFNIRCFSPNALTFR-----YMDVYISEY---NDMTGFSMRQ---GTDKEFRK 217
 DB 174 LSSQDDMMWLR-----IARYKFDYDELIATYVYVGRQISRMKKYITIGRELRIRK 226
 QY 218 RLPMY 222
 DB 227 HLDIW 231

RESULT 9

hypothetical protein alr3063 [imported] - Nostoc sp. (strain PCC 7120)

Species: Nostoc sp.

Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AM2188

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AM2188

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-330 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA874762.1; PID:q17132157; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3063

Query Match

Best Local Similarity 26.9%; Pred. No. 6e-11; Length 330;

Matches 56; Conservative 48; Mismatches 81; Indels 23; Gaps 5;

QY 5 VESIIIPFNAAVTLQACISIVGQTYREVEVYLVDSGSDTRTLDIANSFRELGSRLVY 64

DB 6 LISVIIPYNNVLTIRKETIESYKOTFTDMEIIVINDGSTDGTPELIGIKDE---RLKI 62

QY 65 HSGPDGPRYDAMNRCGVATGEMVLEFLGADDTLYEPTTLAQVAALFGDHAASHLYV---- 120

DB 63 FNYKNGGLPVAKNRGILASGEFLADDD-LMAVDKLEMLKALQHPGEGVAYSWTC 121

QY 121 -----GDVWRSSTKSRHAGPFDDLRLLEFETNLCH--OSIYRRELFPGIGPYNRVYR 171

DB 122 FMDVDEGEPAVYALPSSQYSGFYGNILVSDFTHSGSNTLIRKALNSVGEFDEMLKS 181

QY 172 WADMDFNIRCFSPNALTFRYMDVYSEY 199

DB 182 CEDMDYMLR-----LAWHMDFTIVPEY 203

RESULT 10

hypothetical protein RV1514C - Mycobacterium tuberculosis (strain H37Rv)

Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: E70714

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:96295987; PMID:9634230

A:Accession: E70714
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-262 <COL>
 A:Cross-references: GB:479701; GB:AL123456; NID:q3261635; PIDN:CAB02028.1; PID:626414
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: RV1514C

Query Match

Best Local Similarity 32.2%; Pred. No. 6.8e-11; Length 262;

Matches 68; Conservative 36; Mismatches 90; Indels 17; Gaps 9;

QY 2 TAPVESIIIPFNAAVTLQACISIVGQTYR-EVEVYLVDSGSDTRTLDIANSFRELGS 60

DB 3 SAPTVSVITISFNDLDGLQRTKSVRAQRYRGRLEIHVIDGSGDDVAVYLSGCEPFA- 61

QY 61 RLVHSGPDDGPRYDAMNRCGVATGEMVLEFLGADDTLYEPTTLAQ-VAALFGDHAASHLY 119

DB 62 --YWGSEPDGGRYDAMNRCGVATGEMVLEFLGADDTLYEPTTLAQ-VAALFGDHAASHLY 118

QY 120 YG---DVWRSSTKSRHAGPFDDLRLLEFETNLCH--OSIYRRELFPGIGPYNRVYR 175

DB 119 WGRMDRLVGLDVRKPIPSLRKFLAGKQVYRQASFSFSSLYAKTIGYDLPFGIADQ 178

QY 176 DFNIRCFSPNALTFRYMDV--VISEYNDMTG 204

DB 179 EFLIRA-----ALVCEVTIRVCVCEP-DTIG 204

RESULT 11

hypothetical protein al10914 [imported] - Nostoc sp. (strain PCC 7120)

Species: Nostoc sp.

Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AG1920

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AG1920

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-337 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA872871.1; PID:q17130260; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: al10914

Query Match

Best Local Similarity 14.0%; Score 196.5; DB 2; Length 337;

Matches 61; Conservative 39; Mismatches 80; Indels 37; Gaps 8;

QY 7 SIIIPFNAAVTLQACISIVGQTYREVEVYLVDSGSDTRTLDIANSFRELGSRLVY 66

DB 14 SVIIPFENKEMIRKRAIESCLTQTYADVELIYIDGSTDSLEIKTF-----GNEITWRS 69

QY 67 GPDDGPRYDAMNRCGVATGEMVLEFLGADDTLYEPTTLAQVAALFGDHAASHLYGDVYR 126

DB 70 YPOOGGNHARNRNGFELSOGREYIQLYADADYL-PEKIEKOVNLF-ETTGADVAYGD---- 123

QY 127 SIKSRHAGPFDDLRLLEFETNLCH--OSIYRRELFPGIGPYNRVYR 170

DB 124 WRQRRL-PDSSSEFLKIRISGVQADILALNMMWALAALMYKRSVENSDDRMETPL 182

QY 171 WADMDFNIRCFSPNALTFRYMDVYSEY 197

DB 183 AAGDRFFLSVYVWNGAKVAVQPCVAVYRKYGSVYTS 219

RESULT 12

G95948

probable glycosyltransferase protein SMD21189 [imported] - *Sinorhizobium meliloti* (strain C:Species: Sinorhizobium meliloti)

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: G95948

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernat

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1.683-kb PSYMB megaplasmid from the N2-fixing end

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G95948

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <KUR>

A:Cross-references: GB:AL591985; PIDN:CA049255.1; PID:g15140741; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid PSYMB

R:Galber, F.; Finan, T.M.; Long, S.R.; Punter, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowle, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lejandre,

heault, P.; Vandenhol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMD21189

A:Genome: plasmid

Query Match 14.0%; Score 196.5; DB 2; Length 367;

Best Local Similarity 30.7%; Pred. No. 3.7e-10;

Matches 54; Conservative 33; Mismatches 78; Indels 11; Gaps 4;

QY 1 MTAPEVSIITPTNAATLQACGSIYQYREVEVVLVDGSDTRTDIANSPFELG 60

DB 1 MPEVSVIIPVNAEPYIAAIESVLRQYERIEVIAIDGSDTRSDIERYR-KSDS 59

QY 61 RLVAHSGPDGPDAMRGVATGEVWLFGLGADDTLEPTTLAQAFLGDHMA5---- 116

DB 60 RVLSISENGLATLNEGGLAKGELIARDADDIAPSLSKQVLFSPRLASGT 119

QY 117 --HLVYGDVVMRSTKSRHAGFFDLRL-LFETNLCHQSIYFRRELDGIGPYNLR 169

DB 120 GIDMLIGNRIIRGKPNPIYRGSILRIISMFTTIFMHSIVYNNRV---IPEMLRY 172

RESULT 13

glycosyltransferase [imported] - *Clostridium acetobutylicum*

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: A97168

R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: A97168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80132.1; PID:g15025168; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2174

Query Match 13.7%; Score 192; DB 2; Length 336;

Best Local Similarity 23.6%; Pred. No. 8.5e-10;

Matches 65; Conservative 57; Mismatches 113; Indels 40; Gaps 10;

QY 4 PVESIIPTFNAATLQACGSIYQYREVEVVLVDGSDTRTDIANSPFELGSLV 63

DB 5 PKVSIWVYVYSEKYLAEKISIIDQYINDEFIIVDGSIDESYNIISVANK-DNKII 63

QY 64 VHSQDDGPYDAMRGVATGEVWLFGLGADDTLEPTTLAQAFLGDHMA 115

DB 64 VISRHRGLVSLNGINIRAKYIARDADDISINNIREKQFELFLKDDVILIGTRIE 123

QY 116 SHLVYGDVVMRST--KSRHAGFFDLRL-LFETN--LCQSIYFRRELDGIGPYNLR 168

DB 124 A---FGDIDKQKTYNSAFSISKFDSQNEQVFLNSCAIPHSVAFKKSIVKLRGKRE 180

QY 169 YRVADNMFNTRCFSPNPLITPYMDVISEYN-----DMTGFSGMG-- 210

DB 181 YDFAEDYDMLMLAINGKYR-MDECIKYVHNKSKTAVEMFNPKEVYTKAKIDYI 239

QY 211 TDKEPRKRLPYEYVVA--GMECRRLMLAFKRN 243

DB 240 NDINRKKVDYILMGASTGKLVKRYVESTIDKEN 274

RESULT 14

AE2160

glycosyltransferase alr2836 [imported] - *Nostoc* sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AE2160

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriig

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE2160

A:Status: preliminary

A:Residues: 1-322 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA874535.1; PID:g17131930; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr2836

Query Match 13.6%; Score 190.5; DB 2; Length 322;

Best Local Similarity 29.1%; Pred. No. 1.1e-09;

Matches 53; Conservative 39; Mismatches 75; Indels 15; Gaps 7;

QY 7 SIIITPTFNAATLQACGSIYQYREVEVVLVDGSDTRTDIANSPFELGSLV 66

DB 4 SVLISNYTARYLSRAINSVLAQTHSDIIVYDDSTNSRDTYQLQEQAPDKIPF 63

QY 67 GPDGPDYDAMRGVATGEVWLFGLGADDTLEPTTLAQAFLGDHMA5SHLVYGDVVM 125

DB 64 QANOGGAGFAGNAGPAATGEVAFDLADAD-VKPKRLQRIYEFQISDVYGVVHHLDTID 122

QY 126 RSTK---SRHAGP---FDLRLFETN--LCQSIYFRRELDGIGPYNLR 172

DB 123 GNDKRTDQASTGCPKLSIEDLAVIILQTNAMCFPPISGLAYREVELEKFFIDPVKRWIM 182

QY 173 AD 174

DB 183 AD 184

RESULT 15

AD2189

hypothetical protein alr3067 [imported] - *Nostoc* sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AD2189

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriig

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD2189

A:Status: preliminary

Molecule type: DNA

A:Residues: 1-323 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA074766.1; PID:917132161; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3067

Query Match 13 6%; Score 190.5; DB 2; Length 323;

Best Local Similarity 22.2%; Pred. No. 1.1e-09; Mismatches 69; Indels 73; Gaps 6;

Matches 53; Conservative 44; Mismatches 69; Indels 73; Gaps 6;

QY 4 PVFSIIIPFNAAVTLQACISIVGQYREVEVLVDGSGSTRTLDIANSFRPELGSRLV 63

DB 11 PLISVITIPYNGEKTIETIASVQHOTFLDIETIYINDGSTDNTFELVNRNQ---DNRLK 67

QY 64 VHSQPDGPGYDAMNRGVATGEMWLELGAD-----TL 97

DB 68 IFSYENGGLPVARNRGITFHAVGQFIADADDLMTDKLELQFALQYEPENGLAYSMTY 127

QY 98 YEPTTLAQVAA-----FLGDHAASHLYGDVYMRSTKSRHAGPEDLDRLLEFNLCHQSI 152

DB 128 YKFANEADSYADESNSFAGD-----VVAELLIKNFLQNGSNP----- 164

QY 153 FYRRELFQDGIQPPYNLRYRVAMDMDFNRCFS-----NPALITRYMDVY 195

DB 165 LIRRAAIDSVGLFDPLTKCEDMDFTLRLAQKQFALVKKAOITTYROSPTAMTSKLDM 223

Search completed: April 17, 2003, 17:28:55
Job time : 56 secs

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:20:40 ; Search time 75 Seconds

(without alignments)
472.596 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401

Sequence: 1 MTAAPVFSLIIPTFNAAVTLQ.....ALRRLIRKAVKERSNAEP 266

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

arched: 908470 seqs, 133250620 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database:

A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 1401 | 100.0 | 266 | 18 | AAW21774 |
| 2 | 1401 | 100.0 | 266 | 18 | AAW21775 |
| 3 | 1401 | 100.0 | 266 | 18 | AAW21783 |
| 4 | 298 | 21.3 | 297 | 20 | AAW88309 |
| 5 | 205.5 | 14.7 | 298 | 22 | AAW6313 |
| 6 | 202.5 | 14.5 | 262 | 18 | AAW21779 |
| 7 | 201.5 | 14.4 | 321 | 23 | ABP36805 |
| 8 | 190 | 13.6 | 248 | 20 | AAW88314 |
| 9 | 186.5 | 13.3 | 316 | 21 | AAW54072 |
| 10 | 186.5 | 13.3 | 316 | 21 | AAW43774 |

| | | | | | |
|----|-------|------|-----|----|----------|
| 11 | 186 | 13.3 | 277 | 18 | AAW21767 |
| 12 | 184.5 | 12.9 | 327 | 22 | AAW21767 |
| 13 | 180.5 | 12.9 | 327 | 23 | ABP26804 |
| 14 | 179 | 12.8 | 278 | 21 | AAW68976 |
| 15 | 175 | 12.5 | 334 | 21 | AAW68963 |
| 16 | 174.5 | 12.5 | 333 | 23 | ABW53519 |
| 17 | 174.5 | 12.5 | 623 | 23 | ABW48553 |
| 18 | 172 | 12.3 | 358 | 22 | AAU34218 |
| 19 | 172 | 12.3 | 573 | 22 | AAU34218 |
| 20 | 167.5 | 12.0 | 295 | 23 | ABP27408 |
| 21 | 165.5 | 11.8 | 336 | 21 | AAW54098 |
| 22 | 165.5 | 11.8 | 336 | 21 | AAW54098 |
| 23 | 165 | 11.6 | 251 | 22 | AAW68962 |
| 24 | 163 | 11.6 | 332 | 21 | AAW68962 |
| 25 | 162.5 | 11.6 | 274 | 22 | AAW68962 |
| 26 | 161.5 | 11.5 | 346 | 23 | AAU72923 |
| 27 | 161 | 11.5 | 324 | 18 | AAW14078 |
| 28 | 161 | 11.5 | 324 | 18 | AAW22177 |
| 29 | 160.5 | 11.5 | 322 | 21 | AAW54071 |
| 30 | 160.5 | 11.5 | 322 | 21 | AAW54071 |
| 31 | 158.5 | 11.3 | 327 | 21 | AAW54095 |
| 32 | 158.5 | 11.3 | 327 | 21 | AAW54095 |
| 33 | 158 | 11.3 | 348 | 17 | AAW91311 |
| 34 | 158 | 11.3 | 348 | 17 | AAW06576 |
| 35 | 157.5 | 11.2 | 706 | 22 | AAU33454 |
| 36 | 157.5 | 11.2 | 715 | 22 | AAU33454 |
| 37 | 157 | 11.2 | 327 | 23 | ABW48565 |
| 38 | 156.5 | 11.2 | 270 | 21 | AAW97203 |
| 39 | 153 | 10.9 | 322 | 21 | AAW68974 |
| 40 | 152.5 | 10.9 | 328 | 21 | AAW81720 |
| 41 | 152 | 10.8 | 322 | 21 | AAW81720 |
| 42 | 149.5 | 10.7 | 972 | 20 | AAW06212 |
| 43 | 149.5 | 10.7 | 972 | 20 | AAW43099 |
| 44 | 148.5 | 10.6 | 963 | 23 | ABW54234 |
| 45 | 147.5 | 10.5 | 965 | 21 | AAW96213 |

ALIGNMENTS

RESULT 1
ID AAW21774 standard; Protein; 266 AA.
XX AAW21774;
AC AAW21774;
XX 11-MAR-1998 (first entry)
DT 11-MAR-1998 (first entry)
XX Protein encoded by ORF F of GS region in *M. avium* subspecies *silvaticum*.
DE GS; pathogenesis island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease.
XX
XX Mycobacterium avium subspecies silvaticum.
OS WO9723624-A2.
PN 03-JUL-1997.
PD 23-DEC-1996; 96WO-GB03221.
XX 21-DEC-1995; 95GB-0026178.
PR (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX Doran T, Ford J, Hermon-Taylor J, Loughlin M, Millar D;
PI Sumar N, Tizard M;
XX WPI; 1997-351061/32.
DR N-PSDB; AAT74469.
XX New isolated pathogenicity island from mycobacteria - used to

Protein encoded by
EpsA, Lactococcus
Streptococcus poly
CpsIX protein whic
CpsIX protein whic
Lactococcus lactis
Listeria monocytog
Staphylococcus aur
Staphylococcus aur
Streptococcus poly
Enzyme EpsA Invol
Amino acid sequenc
C glutamicum prote
CpsIX protein whic
C glutamicum prote
Neisseria meningit
S. thermophilus exo
S. thermophilus exo
Enzyme Eps7 which
Amino acid sequenc
Enzyme Eps7 Invol
Amino acid sequenc
N. gonorrhoeae gly
Neisseria polyglyc
Enterococcus faeca
Listeria monocytog
Campylobacter jej
CpsII protein whic
Streptococcus pneu
CpsIX protein whic
Pasteurella multoc
P. multocida hyalu
Lactococcus lactis
P. multocida chond

PT develop products for detection, diagnosis, prevention and treatment
of mycobacteria infections

Claim 1; Page 50; 62pp; English.

CC The present sequence represents a protein encoded by open reading frame
CC (ORF) F of a novel polynucleotide sequence designated "GS". GS is a
CC pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb
CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
CC ORFs, and also the transmissible element, encode proteins which may be
CC linked to pathogenecity, such as providing receptors for cellular
CC recognition. GS was discovered and characterised using differential DNA
CC analysis technology. It is found within Mycobacterium paratuberculosis
CC and it has also been identified in Mycobacterium avium subspecies
CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
CC the intestine and Crohn's disease in humans. The protein products of the
CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
CC treating or preventing mycobacterial disease. In particular they can be
CC used as vaccines for inflammatory diseases such as Crohn's disease or
CC sarcoidosis in humans or John's disease in animals.

Sequence 266 AA;

Query Match 100.0%; Score 1401; DB 18; Length 266;
Best Local Similarity 100.0%; Pred. No. 7.3e-144;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTAPEFSIIIPFNAAVTLQACISIVGQTYREVEVVLVDGSGTDRITDIANSFRELGS 60
DB 1 MTAPEFSIIIPFNAAVTLQACISIVGQTYREVEVVLVDGSGTDRITDIANSFRELGS 60
QY 61 RLTVHSGPDDGPDYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120
DB 61 RLTVHSGPDDGPDYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120
QY 121 GDVYMRSTKSRHAGFPDLDRLLEFETNLCHOSIFYRRELDGIGPYNLRYRWADMPNIR 180
DB 121 GDVYMRSTKSRHAGFPDLDRLLEFETNLCHOSIFYRRELDGIGPYNLRYRWADMPNIR 180
QY 181 CFSNPALITRYMDVYISEYNDMTGFSMROGTDKEFRRLPMYFWAGWETCRMLAFLKD 240
DB 181 CFSNPALITRYMDVYISEYNDMTGFSMROGTDKEFRRLPMYFWAGWETCRMLAFLKD 240
QY 241 KENRRALRLRLIRYKAVSKERSAEP 266
DB 241 KENRRALRLRLIRYKAVSKERSAEP 266

RESULT 2

AAW21775 standard; Protein; 266 AA.

AAW21775;

11-MAR-1998 (first entry)

Protein encoded by ORF F of the GS region in M. paratuberculosis.

GS; pathogenecity island; pathogenic protein; mycobacterial disease;
cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
vaccine; inflammatory disease; sarcoidosis; John's disease.

Mycobacterium paratuberculosis.

W09723624-A2.

03-JUL-1997.

23-DEC-1996; 96W0-GB03221.

21-DEC-1995; 95GB-0026178.

(SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX Doran T, Ford J, Hermon-Taylor J, Loughlin M, Millar D;
PI Sumar N, Tizard M;
XX
DR WPI: 1997-351061/32.
DR N-PSDB: AAT74470.

PT New isolated pathogenecity island from mycobacteria - used to
develop products for detection, diagnosis, prevention and treatment
of mycobacteria infections

Claim 1; Page 51; 62pp; English.

CC The present sequence represents a protein encoded by open reading frame
CC (ORF) F of a novel polynucleotide sequence designated "GS". GS is a
CC pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb
CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
CC ORFs, and also the transmissible element, encode proteins which may be
CC linked to pathogenecity, such as providing receptors for cellular
CC recognition. GS was discovered and characterised using differential DNA
CC analysis technology. It is found within Mycobacterium paratuberculosis
CC and it has also been identified in Mycobacterium avium subspecies
CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
CC the intestine and Crohn's disease in humans. The protein products of the
CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
CC treating or preventing mycobacterial disease. In particular they can be
CC used as vaccines for inflammatory diseases such as Crohn's disease or
CC sarcoidosis in humans or John's disease in animals.

Sequence 266 AA;

Query Match 100.0%; Score 1401; DB 18; Length 266;
Best Local Similarity 100.0%; Pred. No. 7.3e-144;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTAPEFSIIIPFNAAVTLQACISIVGQTYREVEVVLVDGSGTDRITDIANSFRELGS 60
DB 1 MTAPEFSIIIPFNAAVTLQACISIVGQTYREVEVVLVDGSGTDRITDIANSFRELGS 60
QY 61 RLTVHSGPDDGPDYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120
DB 61 RLTVHSGPDDGPDYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120
QY 121 GDVYMRSTKSRHAGFPDLDRLLEFETNLCHOSIFYRRELDGIGPYNLRYRWADMPNIR 180
DB 121 GDVYMRSTKSRHAGFPDLDRLLEFETNLCHOSIFYRRELDGIGPYNLRYRWADMPNIR 180
QY 181 CFSNPALITRYMDVYISEYNDMTGFSMROGTDKEFRRLPMYFWAGWETCRMLAFLKD 240
DB 181 CFSNPALITRYMDVYISEYNDMTGFSMROGTDKEFRRLPMYFWAGWETCRMLAFLKD 240
QY 241 KENRRALRLRLIRYKAVSKERSAEP 266
DB 241 KENRRALRLRLIRYKAVSKERSAEP 266

RESULT 3

AAW21783 standard; Protein; 275 AA.

AAW21783;

11-MAR-1998 (first entry)

Protein encoded by ORF F from the GS region of M. tuberculosis.

GS; pathogenecity island; pathogenic protein; mycobacterial disease;
cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
vaccine; inflammatory disease; sarcoidosis; John's disease; ss.

Mycobacterium tuberculosis.

W09723624-A2.

| | |
|----|---|
| AC | AAB96313; |
| XX | |
| DT | 29-OCT-2001 (first entry) |
| XX | |
| DE | Putative glycosyltransferase, involved in cell wall biogenesis #1. |
| XX | |
| KW | Hyperthermophilic archaeon; hyperthermophilic protein. |
| XX | |
| OS | Pyrrococcus abyssi. |
| PN | FR2792651-A1. |
| XX | |
| PD | 27-OCT-2000. |
| PF | |
| PE | 21-APR-1999; 99FR-0005034. |
| PR | 21-APR-1999; 99FR-0005034. |
| XX | |
| PA | (CNRS) CNRS CENT NAT RECH SCI. |
| XX | (IFRE-) IFREMER INST FR RECH EXPL MER. |
| PI | Forterre P., Thierry JC, Prieur D, Dietrich J, Lecompte O; |
| XX | Querellou J, Weissensbach J, Sautin W, Hellig R; |
| DR | WPI; 2001-126236/14. |
| XX | |
| PT | New nucleotide sequences isolated from Pyrococcus abyssi encode |
| PT | proteins useful in industry - |
| XX | |
| PS | Claim 7; Pages 981-982; 1657pp; French. |
| CC | |
| CC | The present invention relates to the genomic sequence of Pyrococcus |
| CC | abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is |
| CC | a hyperthermophilic archaeon, which is isolated from deep-sea |
| CC | hydrothermal vents. The present sequence is one such P. abyssi protein. |
| CC | The proteins of the present invention have various potential industrial |
| CC | uses, since the proteins are stable at very high temperatures, some up to |
| CC | 110 degrees centigrade. |
| CC | Note: This patent is in the same patent family as WO200065062, which |
| CC | contains additional sequences as shown in AAB99132-AB99143, |
| CC | AAH75903-AAH75920 and AAG66436. |
| XX | |
| SQ | Sequence 298 AA: |
| | |
| | Query Match 14.7%; Score 205.5; DB 22; Length 298; |
| | Best Local Similarity 29.0%; Pred. No. 1,le-13; |
| | Matches 71; Conservative 38; Mismatches 99; Indels 37; Gaps 10 |
| OY | 1 MTAPVSIITIPFNAAVTLQACLGSIVGTYYREVEVLVDGSGTDTLTIANSFPBELGS 60 |
| | : : : : : : : : : : : : : : : : : : |
| | 1 MSRPVSVIPIPYNRANMLRRRAIASVLAOKFDLFIVDDASTDWTPEVVEST--EDGR 58 |
| Y7 | 61 ----LLVVHSGDDDDPYDMNMGVGVAATGEWILFLGADDTLYPPTLLAQYAALGSHAAS 116 |
| | : : : : : : : : : : : : : : : : : : |
| Db | 59 IRYIRLKKNKG---SPT-ARNIGIKAKARFTALDDDEMTLPHRLEVOVRKEENLGKER 114 |
| OY | 117 HLIVGDVVARSTRSR-----HAGPFDLRLLFETINLCHOSIFYRRELFDGIGPYNLR 168 |
| | : : : : : : : : : : : : : : : : : : |
| Db | 115 GVYGGGFYYSODGRLKRLPKKHODI-YSHLLKENFTIGSPILLIRRECFFKAAGLFDPR 173 |
| OY | 169 YRVWADMDFENICFSNPALITR----YMDVVISSEY--NDMTGFSMRQ---GTDEKFRK 217 |
| | : : : : : : : : : : : : : : : : : : |
| Db | 174 LSSSDMDMWMRL-----IARYRKDYDELIATAKYVHGKQSFMKKYIPIPERELIRK 226 |
| OY | 218 RLDPK 222 |
| | : : : : : : : : : : : : : : : : : : |
| Db | 227 HLDIW 231 |

```

AC AAM21779;
XX
XX 11-MAR-1998 (first entry)
DT
DE Protein encoded by ORF B from the GS region of M. tuberculosis.
XX
XX GS: pathogenesis island; pathogenic protein; mycobacterial disease;
KM cellular recognition receptor; pathogenic Mycobacteria; Crohn's disease;
XX vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
OS
XX Mycobacterium tuberculosis.
XX
XX MO9723624-AZ.
PN
XX 03-JUL-1997.
PD
XX 23-DEC-1996; 96WO-GB03221.
PF
XX 21-DEC-1995; 95GB-0026178.
PR
XX (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
PA
XX Doran T, Ford J, Hermon-Taylor J, Loughlin M, Millar D;
PI Sumar N, Tizard M;
PI WPI: 1997-351061/32.
XX N-PSDB: AAT74473.
DR
XX New isolated pathogenicity island from mycobacteria - used to
PT develop products for detection, diagnosis, prevention and treatment
PT of mycobacteria infections
XX
XX Claim 1; Page 53; 62pp; English.
XX
XX The present sequence represents the protein encoded by open reading frame
CC (ORF) B, from M. tuberculosis. This ORF B has been found to have homology
CC with ORF B of a novel polynucleotide sequence designated "GS". GS is a
CC pathogenesis island of 8 kb of DNA comprising a core region of 5.75 kb
CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
CC ORFs, and also the transmissible element, encode proteins which may be
CC linked to pathogenesis, such as providing receptors for cellular
CC recognition. GS was discovered and characterised using differential DNA
CC analysis technology. It is found within Mycobacterium paratuberculosis
CC and it has also been identified in Mycobacterium avium subspecies
CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
CC the intestine and Crohn's disease in humans. The protein products of the
CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
CC treating or preventing mycobacterial disease. In particular they can be
CC used as vaccines for inflammatory diseases such as Crohn's disease or
CC sarcoidosis in humans or Johne's disease in animals.
XX
XX Sequence 262 AA:
SO
Query Match 14.5%; Score 202.5; DB 18; Length 262;
Best Local Similarity 32.2%; Pred. No. 1.9e-13;
Matches 68; Conservative 36; Mismatches 90; Indels 17; Gaps 9
DY 2 TAPVSIILIFPNAATLTQAGLSYVGQYR-EVEVVVDGSGTDTLIDANSFRRLGS 60
DB 3 SAPTVSVITISFPNDLDGLDRTVAKSVAQRGRHIEHVIVDGSGDDVAVYLSCEGFSA- 61
OY RLTVHSFGPDGPEVDANRRGVATGVWVFLGADDTLYEPTTLAQ--VAAFILGHASHLV 119
DB 62 --YMDSERPGRGDYDANNQGIHAHSGDLLFLFHSHADFSGEDVYAQAVENTSLSGPPSEL- 118
OY 120 YG---DVNKRSTKSRHAGPPDDLRLFETNLNC-HQSIFYRRELFDSIGYPNLRRYWADM 175
DB 119 WQFGMDRLTGDLRVNRGPFIFFSLKRFIAGQVAPHQASFGSSIVARKIGGYDIDFGIAADQ 178
OY 176 DENIRCFNSPALITRYMDV--VISEYNDMTG 204
DB 179 EELIRA---ALVCEPVYTRCYVCEP-DVTG 204

```


| | | | |
|----|---|---|-------------------------------------|
| XX | ABP26805 | ID | ABP26805 standard; Protein; 321 AA. |
| XX | AC | ABP26805; | |
| XX | DT | 02-JUL-2002 | (first entry) |
| XX | DE | Streptococcus polypeptide SEQ ID NO 2786. | |
| XX | KM | Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae; | |
| XX | KW | group A streptococcus; Streptococcus pyogenes; antibacterial; | |
| XX | KX | antiinflammatory; infection; vaccine; meningitis; gene therapy; | |
| XX | OS | Streptococcus agalactiae. | |
| XX | SN | WO200234771-A2. | |
| XX | MA | 02-MAY-2002. | |
| XX | PF | 29-OCT-2001; 2001WO-GB04789. | |
| XX | PR | 27-OCT-2000; 2000GB-0026333. | |
| XX | PR | 24-NOV-2000; 2000GB-0028727. | |
| XX | PR | 07-MAR-2001; 2001GB-0005640. | |
| XX | PA | (CHIR-) CHIRON SPA. | |
| XX | PA | (GENO-) INST GENOMIC RES. | |
| XX | PI | Telford J, Maignant V, Margarit Ros YI, Grandt G, Fraser C; | |
| XX | PI | Tetella H; | |
| XX | DR | WPI: 2002-352536/38. | |
| XX | DR | N-PSDB: ABN67436. | |
| XX | PT | New Streptococcus protein for the treatment or prevention of infection | |
| XX | PT | or disease caused by Streptococcus bacteria, such as meningitis, and | |
| XX | PT | for detecting a compound that binds to the protein - | |
| XX | PS | Claim 1; Page 3429; 4525pp; English. | |
| CC | CC | The invention relates to a protein (ABP25413-ABP30895) from group B | |
| CC | CC | streptococcus/GBS (Streptococcus agalactiae) or group A Streptococcus/GAS | |
| CC | CC | (Streptococcus Pyogenes), comprising one of 5483 sequences (SI), given in | |
| CC | CC | the specification. The proteins have antibacterial and antiinflammatory | |
| CC | CC | activity. (1), nucleic acids encoding (1), ABN6004-ABN71526 and | |
| CC | CC | antibodies that bind (1) are used in the manufacture of medicaments for | |
| CC | CC | the treatment or prevention of infection or disease caused by | |
| CC | CC | Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. | |
| CC | CC | Nucleic acids encoding (1) are used to detect Streptococcus in a | |
| CC | CC | biological sample. (1) is used to determine whether a compound binds to | |
| CC | CC | (1). A composition comprising (1) or a nucleic acid encoding (1), may be | |
| CC | CC | used as a vaccine or diagnostic composition. The disease caused by | |
| CC | CC | Streptococcus that is prevented or treated may be meningitis. Nucleic | |
| CC | CC | acid encoding (1) may be used to recombinantly produce (1) and may be | |
| CC | CC | used in gene therapy. Antibodies to (1) are used for affinity | |
| CC | CC | chromatography, immunoassays, and distinguishing/identifying | |
| CC | CC | Streptococcus proteins. | |
| XX | XX | Sequence 321 AA: | |
| OY | Query Match | 14.4%; Score 201.5; DB 23; Length 321; | |
| OY | Best Local Similarity | 30.1%; Pred. No. 3,4e-13; | |
| DB | Matches 52; Conservative 33; Mismatches 63; Indels 25; Gaps 6 | | |
| OY | 7 SIITFTNAAVTLQACLSIVGQTREVEVVYVDGGSTDRTDIANSPFELGSRILVHS 66 | | |
| OY | : : : : : : : : : : : : : : | | |
| DB | 5 SIITPYNVQSFLNCIESEVLQAOTYSNMETIIIVNGSTDSNDICD-YSEIDGRIFVH 63 | | |
| OY | 67 GDDDPDYAMANGVGVAIGEWLFLGADDTLYEPTTAAOVAAFLGDHAASHLYTG----- 121 | | |
| OY | : : : : : : : : : : : : : : | | |
| DB | 64 KNNGGISLARNRNGISRATCDITYLLSDSDYLKKEAIEEMVEF-SERYNSEIYLGCYVER 122 | | |

```

0Y 122 -----DVMRSTKSRHAGP-----FDDILRFLFENLGHOSIFVRRLEF 159
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 RQGHINIVLEDEMETTISFVOAIONITVDAIRAIF--TVAHNTL-YKRELF 172

RESULT 8
AAW88314
ID AAW88314 standard; Protein: 248 AA.
AC AAW88314;
XX
XX
XX 26-APR-1999 (first entry)
DE Sugar transferase involved in O157 antigen biosynthesis.
XX O antigen; O157 antigen; sugar transferase; wbdO gene;
KW diarrhoea; haemorrhagic colitis; diagnosis.
XX Escherichia coli.
OS
XX WO9850531-A1.
XX
XX 12-NOV-1998.
XX
XX 01-MAY-1998; 98WO-AU00315.
XX
XX 22-JUL-1997; 97AU-0008162.
PR 01-MAY-1997; 97AU-0006545.
PA (UNSY ) UNIV SYDNEY.
XX
XX
XX Reeves PR, Wang L;
PI
XX WPI: 1999-059669/05.
DR N-PSDB; AAX06749.
XX
XX Nucleic acid molecules specific for bacterial polysaccharide
PT antigens - useful for detecting specific strains in, e.g. food,
PT faeces or patient samples
XX
XX Disclosure: Fig 8; 165pp; English.
XX
XX This is the amino acid sequence of a sugar transferase that is
CC encoded by the wbdO open reading frame of a gene cluster (see
CC AAX06749) involved in the biosynthesis of the Escherichia coli O157
CC O antigen. The use of nucleic acid molecules derived from
CC particular assembly and transport genes, particularly wbd
CC (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC antigen gene clusters improves the specificity of methods for the
CC detection and identification of O antigens, e.g. in testing food-
CC or faecal-derived samples, or samples from patients. The O antigen
CC is a major virulence factor of enteropathogenic E. coli strains
CC that cause diarrhoea and haemorrhagic colitis.
XX
XX
XX Sequence 248 AA;
.S0

Query Match 13.6%; Score 190; DB 20; Length 248;
Best Local Similarity 26.9%; Pred. No. 4,le-12;
Matches 67; Conservative 42; Mismatches 80; Indels 60; Gaps 11;

0Y 7 SIIITFFAAAVYLQAGLSIVGQYREVEVVLVDGSGDTRDLDIANSFPELGSRLVYH-65
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 4 SVITVTYNNABELETLSLSLILKRPETIIIVDGSIDGTNRVYSRLEF-----SANTHYV 59
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

0Y 66 SGPDGDPYDAMNRGVAVGEMVFLGADDTLYEPPTLQAQVAFLGDHAASHLVYGDVYM 125
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 60 YEKDEGIYDAMNKGRLAKGDLIHNLNGDSV-----IGD-----YKNI--99
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

0Y 126 RSTSRHAGPFDLDDL-----FETNLCHOSITRYRELPGIGTIPYMLRIRWADMDPQN 178
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 100 KPECLIKVGLPEENDLGLFSSITHSNTGYCHQGVIFPKNHS-----YDLRYICADYKLI 155
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

QY 179 IRCSPNALLIRIVDYISEY--NDWICFSMRGG--TDNEFRRLPMFWAGMETCRM 234
 Db 156 QGVPEPES---LRSLSLTISGIVKXDMGVSSKKRLRLKRELAK----- 195
 QY 235 LAFLKDKEN 243
 Db 196 IMPERKKEN 204

RESULT 9
 AAY54072 ID AAY54072 standard; Protein: 316 AA.
 AC AAY54072;
 XX 27-MAR-2000 (first entry)
 DE Enzyme EPS8 which is involved in exopolysaccharide biosynthesis.
 KM Exopolysaccharide; EPS; ESP enzyme; EPS1; EPS2; EPS3; EPS4; EPS5;
 KM EPS6; EPS7; EPS8; EPS9; EPS10; Streptococcus thermophilus strain Sfi139;
 KM activated D-galactose pyranose; saccharide; beta-1,4-galactosyltransferase;
 KM transporter; food; fermented milk product; yoghurt; cheese;
 KM flavour stability; organoleptic property.
 XX Streptococcus thermophilus.
 OS WO9962316-A2.
 XX 09-DEC-1999.
 XX 22-APR-1999; 99WO-EP02841.
 XX 22-APR-1998; 98EP-0201310.
 XX 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX (NEST) SOC PROD NESTLE SA.
 XX Stingele F, Germond JE, Lamothe G;
 DR WPI: 2000-097267/08.
 DR N-PSDB; AA245256, AAY54072, AAY54073, AAY54074.
 PT New recombinant enzymes for synthesis of exopolysaccharides,
 PT particularly in lactic acid bacteria, for improving properties of
 PT fermented milk products -
 XX Claim 3; Page 105-106; 162pp; French.

AAY54065-74 represent enzymes involved in the biosynthesis of
 exopolysaccharides (EPS). These enzymes are designated EPS1-EPS10, and
 are encoded by open reading frames. eps1-eps10. The enzymes are isolated
 from Streptococcus thermophilus strain Sfi139. The proteins are used
 in a method for the synthesis of EPS, which includes at least one step
 of forming a bond (alpha or beta-isomer) between C-1 (carrying the
 reducing aldehyde function, of an activated D-galactose pyranose), and
 a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis
 of EPS occurs with, in each step, addition of a new sugar unit, through
 its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar
 unit, present at the end of a chain of sugar residues bonded to the
 primer. EPS1 to EPS4 are used to elongate the EPS chain, and to regulate
 EPS production, EPS5 creates new bonds between saccharides, EPS6 is used
 in the biosynthesis of EPS, EPS7 and EPS8 are beta-1,4-galactosyltransferases,
 CC EPS9 is a transporter of repetitive units, and EPS10 catalyses the
 CC conversion of a pyranose form of a beta-D-galactose to the furanose
 CC form. The EPS enzyme are used to improve properties of foods,
 CC particularly fermented milk products such as yoghurt and cheese,
 CC e.g. their organoleptic properties and flavour stability.
 XX Sequence 316 AA;
 SQ Query Match 13.3%; Score 186.5; DB 21; Length 316;

Best local Similarity 28.9%; Pred. No. 1.4e-11;
 Matches 59; Conservative 42; Mismatches 74; Indels 29; Gaps 9;

QY 4 PVFSIIIPTEFNAVTLQACIGSIVGQTYREVEVVLVDGSTDRLDIANSRPELSRLV 63
 Db 3 PLISIIIVPYNNEKYIRFQIESIIAQTYYNINIEIYINDGSTDSLAVISLCSHINIKV 62
 QY 64 VHSQPDGPPYDAMNRGVATGEMVLEFLGADDTLYEP---TTIAQVAAFLGDHAASHLYY 120
 Db 63 INQ-KNOGLSVARNTGIDAATGKXIAFVADADKRI-KPDEFVSLXQIA---DKTGADIVR 116
 QY 121 GDV-----VMSRKSRAQFPDDRLLEFNILCH-QSIFYRRLPGICIPYNL 167
 Db 117 GSFREDENGNIPIKGVWPDFNVPTNYGTIVLDQFL-SSNISFVMSIYR---LDFINSNHI 172
 QY 168 RYR---VMADMPENIRCFSPNALI 188
 Db 173 RFTPGILFEDADFTIRAYIAKLIV 196

RESULT 10
 AAY43774 ID AAY43774 standard; Protein: 316 AA.
 AC AAY43774;
 XX 11-FEB-2000 (first entry)
 DE Amino acid sequence of eps8 of Streptococcus thermophilus Sfi139.
 XX eps operon; Streptococcus thermophilus Sfi139; enzyme; eps1; eps2;
 KM eps3; eps4; eps5; eps6; eps7; eps8; eps9; eps10;
 KM exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;
 KM probiotic; foodstuff; organoleptic quality; flavour;
 KM lactic acid bacteria; acidified milk product; yoghurt; cheese.
 XX Streptococcus thermophilus.
 OS WO9954475-A2.
 XX 28-OCT-1999.
 XX 22-APR-1999; 99WO-EP03011.
 XX 22-APR-1998; 98EP-0201310.
 XX 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX (NEST) SOC PROD NESTLE SA.
 XX Stingele F, Germond JE, Lamothe G;
 DR WPI: 2000-013255/01.
 DR N-PSDB; AA230355, AAY43774, AAY43775, AAY43776;
 PT New recombinant enzymes for biosynthesis of exopolysaccharides having
 PT e.g. antitumour or probiotic properties or useful in fermented milk
 PT products -
 XX Claim 3; Page 106-107; 163pp; French.

AAY43767-76 represent the proteins encoded by the eps operon of
 Streptococcus thermophilus Sfi139. The operon contains 10 open reading
 frames, and encodes enzymes (eps1, eps2, eps3, eps4, eps5, eps6, eps7,
 CC eps8, eps9 and eps10) that are involved in the biosynthesis of
 CC exopolysaccharides (EPS). The enzymes catalyse the formation of
 CC specific intersugar bonds. The enzymes catalyse a process that includes
 CC at least one step of forming a bond (in alpha or beta anomeric form)
 CC between C1, carrying the reducing aldehyde group of an activated D-Galp
 CC (galactose in pyranose form), and a phosphate group on a lipophilic or
 CC proteinaceous primer. The enzymes are used to produce EPS that have
 CC antitumour or probiotic properties or are used in foodstuffs to improve
 CC organoleptic qualities and flavour. When expressed by lactic acid

PI Telford J, Maslignani V, Margarit Ros YI, Grandi G, Fraser C.

complement-mediated crossphagocytosis: serotype-specific detection;
capsular component; antigen; regulation; chain length determination;
KW

QY 236 -APLKDKENRRLLAL 248
: : : : :
Db 283 IVYKQOLKONKRRLLAL 296

Search completed: April 17, 2003, 17:26:56
Job time : 77 secs